COVID-19 and controlled distance in the state of Rio Grande do Sul: an analysis of coping measures

DOI: 10.46932/sfjdv3n3-066

Received in: March 22nd, 2022
Accepted in: April 21st, 2022

Eduardo Damiani Pavin
Master's student in Business Administration at the Graduate Program in Business Administration at Universidade de Passo Fundo (PPGAdm/UPF)
Institution: Universidade de Passo Fundo (PPGAdm/UPF)
Address: Fundação Universidade de Passo Fundo, BR 285 Km 292,7, Campus I, Bairro São José, CEP: 99052-900, Passo Fundo – RS, Brasil
E-mail: 158342@upf.br

Luis Henrique Botton
Master's student in Business Administration at the Graduate Program in Business Administration at Universidade de Passo Fundo (PPGAdm/UPF)
Institution: Universidade de Passo Fundo (PPGAdm/UPF)
Address: Fundação Universidade de Passo Fundo, BR 285 Km 292,7, Campus I, Bairro São José, CEP: 99052-900, Passo Fundo – RS, Brasil
E-mail: 158349@upf.br

Geraldo Edmundo Silva Junior
Doctor in Economics from the Federal University of Rio Grande do Sul (UFRGS)
Institution: Federal University of Rio Grande do Sul (UFRGS)
Address: Universidade Federal de São Carlos, Campus Sorocaba, Rodovia João Leme dos Santos, Km 110, 264, Itinga, CEP: 18052-780, Sorocaba - SP, Brasil
E-mail: gedmundos@ufscar.br

Andre da Silva Pereira
Doctor in Economics from the Federal University of Rio Grande do Sul (UFRGS)
Institution: Federal University of Rio Grande do Sul (UFRGS)
Address: Fundação Universidade de Passo Fundo, BR 285 Km 292,7, Campus I, Bairro São José, CEP: 99052-900, Passo Fundo – RS, Brasil
E-mail: andresp@upf.br

Geizi Cássia Bettin do Amarante
Master in Business Administration from the Graduate Program in Business Administration at the University of Passo Fundo (PPGAdm/UPF)
Institution: Universidade de Passo Fundo (PPGAdm/UPF)
Address: Fundação Universidade de Passo Fundo, BR 285 Km 292,7, Campus I, Bairro São José, CEP: 99052-900, Passo Fundo – RS, Brasil
E-mail: 126993@upf.br

ABSTRACT
The Coronavirus pandemic (COVID-19) started in Wuhan, China, in December 2019. It was disseminated to the globalized world, affecting it in a multifaceted way in its economic, social, political, and, mainly, public health aspects. Therefore, studying the problem from identifying its pandemic characteristics and its main determinants became relevant from the beginning. This paper used the Machine Learning

approach to identify the main variables that reveal policy changes' effect. Then, the methodological approaches revealed the use of LASSO to Linear Regression; LASSO for Multinomial Logistic Regression, Elastic Net Regression; Regression Tree, and Random Forest for the description of the pandemic in the Rio Grande do Sul. The results showed high accuracy in the identification of the relevant variables with a low error rate. Finally, it concluded that the three possible states of recovered, death, and recovering stage described using the machine learning methodology with a high degree of accuracy, attesting the adequacy of the measures adopted in the Rio Grande do Sul.

Keywords: COVID-19, coronavirus, government policy, public policy, state and local government: health.

1 INTRODUCTION

The pandemics, such as Coronavirus (COVID-19), affect a relatively large number of people and impose new social rules and habits for the world population. Furthermore, the social distance was adopted in Brazil as a measure of preventing the spread of COVID-19, which can have economic and psychosocial consequences (DUARTE *et al*., 2020). The coronavirus was classified as a pandemic for the World Health Organization (WHO) on March 11, 2020, and had its start in Wuhan, China, in December 2019, from which it has taken on a vast dimension, spreading out in a pandemic way for the whole world. According to WHO, on January 7, 2020 the Chinese authorities confirmed the identification of a new coronavirus, as a factor of pneumonia.

In Brazil, the COVID-19 is progressing exponentially. Although the disease has been spread rapidly in the major capitals, where the incidence of cases is high, the cases of COVID-19 increase in the smaller cities and poorer communities (DANTAS *et al*., 2020). More than three-quarters of the cases are in the south and southeast regions of Brazil, which are more densely populated, including elderly people, and with tropical and subtropical climates. Additionally, Dantas *et al*. (2020) emphasizes that the economic burden that sustained distance can impose is potentially catastrophic in Brazil and other developing nations.

Through the ability to spread rapidly, according to Sousa *et al*. (2020), the transmission of the virus occurs from one person sick to another through the droplets of saliva, sneeze or cough, and also through close contact by shaking hands or even handling contaminated objects or surfaces. In consideration of the difficulties to contain the progress of cases and deaths, governments and institutions are adopting models for implementing controlled distancing.

According to the WHO strategic plan, the alternatives to prevent the spread of the virus are in limiting person-to-person transmission, including reducing secondary infections between close contacts and health care workers, and preventing amplification of transmission. In addition, early identification, isolation, and treatment of patients is of interest, including the provision of optimal care for infected patients. Regarding the socio economic aspect, it is worth pointing out that public policies are also needed at the local level, according to the various regional realities (PINTO; CORONEL; MULLER, 2020). Dantas *et al*., (2020) they also point out that the consequences of a pandemic in a country where income inequality is high and the national economy is fragile, such as Brazil, can lead to multifaceted situations, such as the decline of the population's mental health, the collapse of the health system, economic, social and moral crisis.

Considering the global dissemination of the disease, on January 22, 2020, the Public Health Emergency Operations Center for the new Coronavirus was activated (COE-nCoV), whereas on february 26, 2020 the Ministry of Health confirmed the first case of Coronavirus in the country, and from that date, the transmission was significantly increasing, reaching 1,577,004 cases on July 4 and more than 64,000 confirmed deaths from the disease, representing a lethality rate of 4.1%. In Rio Grande do Sul, the focus state of this research, the State Health Department of Rio Grande do Sul (SES/RS) registered the first case of infection on March 10, reaching 31,619 cases of the disease on July 4, with 715 deaths, a lethality rate of 2.3%, 1.8% less than the national average.

The SES/RS published the strategies of confronting Coronavirus, which include, prepare and train the medical infrastructure to attend to the sick, mitigate the economic losses, reorganize the offer of services to citizens, and also regulate social coexistence to avoid contagion. To fulfill this objective, the State Government must act systematically in the axes of health, economics, social and security.

Considering the current pandemic scenario, this article seeks to identify the impacts of the sanitary measures taken by the state of Rio Grande do Sul and their potential effects on the development of the number of confirmed cases, deaths, and the number of patients in treatment status. Based on such measures aimed at confronting the Coronavirus (COVID-19), the objective of the present work is to analyze factors that explain the evolution of COVID-19 cases in the state. For this, the two main state decrees originally issued on 03/16/2020, 03/19/2020, and 05/10/2020 were considered. These decrees refer to the prohibition of non-essential public and private activities and services, the closing of commercial centers, and specific measures for essential services to attend to the population.

The international empirical literature about COVID-19 makes it possible to divide the contributions into three distinct groups, namely:

(i) papers linked to the policy implications of Coronavirus contagion models such as Avery et al. (2020) and Atkenson (2020);

(ii) related research on the available data set and the effects of measures such as Stock (2020), Courtemanche et al. (2020), Hsiang et al. (2020), and Chernozhuzov, Kasahara, and Schrimpf (2021), although the approaches are more counterfactual; and,

(iii) studies focused on the use of artificial intelligence and machine learning such as Ardabili et al. (2020) and Yadav, Perumal, and Srinivas (2020).

When considering the decrees on 03/16/2020, 03/19/2020 and 05/10/2020, a robust approach was chosen that corresponds to the use of machine learning, having as an essential characteristic the use of Lasso (least absolute shrinkage and selection operator) from linear regression, for multinomial logistic regression, elastic net regression, regression tree and random forest.

It should be emphasized that the focus of the present work is not similar to the counterfactualism present in the works of Atkenson (2020), Avery *et al*. (2020), Courtemanche *et al*. (2020), Hsiang *et al*. (2020), Stock (2020) e Chernozhuzov, Kasahara e Schrimpf (2021), but in results based on the causal relationship between the evolution of the patients status, disaggregated into recovered, death, or recovering, and the variables sex, age group, race or color, region, and testing criteria. Furthermore, to corroborate this study, the set of covariates was used, formed by the variables IBGE code of the location, the date of confirmation of the test result, the dates of symptoms, the date of progression to the result, the date of hospitalization, the date of death, and the date of estimated progression. The proposed perspective is based on the factual results and not on the counterfactualism present in the literature.

To achieve that purpose, the present work follow distributed in this manner: the present introduction as a first section; in the second part a theoretical framework is presented as a background for the development of the work; in a third section, the machine learning methodology for the regressions used is presented; in the fourth section the results and discussion are presented; finally, in a last section, the concluding remarks are presented.

2 THEORETICAL FRAMEWORK

2.1 THE GLOBAL PANDEMIC LANDSCAPE AND EXPONENTIAL GROWTH

Considering the dimension and circulation capacity of the virus, it is necessary to adopt strategies to contain discrimination. In this scenario, the WHO has developed a document with strategies and plans to provide support to all countries to prepare for the confrontation with COVID-19. Therefore, the organization advised that countries should urgently take the necessary measures to decrease the spread and prevent their health systems from being overburdened, thus ensuring social welfare (CVETKOVI'C et. al, 2020).

According to Pinto, Coronel and Müller (2020), the new Coronavirus (COVID-19) is a public health emergency that has caused, in a recent context, international concern, the WHO has identified that most infected people will experience mild to moderate respiratory illness and recover without the need for special treatment. Elderly people and people with underlying medical problems such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious complications.

Consequently, Gemicioğlu *et al*. (2020) describes that interpersonal transmission of SARS-CoV-2 has turned the emerging health problem into a global pandemic. Thus, Al-Rohaimi and Al-Otaibi (2020), define that the direct impacts of the pandemic brought human life to a halt, putting global health in a serious situation, highlighting the Coronavirus' ability to circulate.

Besides WHO had prepared a strategic plan for Coronavirus, the debates about these are controversial worldwide. Hale *et al*. (2020) point out that governments have varied substantially in the measures they have adopted and how quickly they have adopted them. This variation has created debate, as policy makers and the population deliberate on what level of response should be pursued and how quickly to implement or reverse them, and as public health experts learn in real time what measures are most or least effective.

Faced with the winding situation that the pandemic is in, Ashraf (2020) states that governments around the world have struggled with emergency actions such as blockades, travel restrictions, testing, quarantine, and economic packages. The main purpose of these actions was to ensure social distancing between people to contain the spread of the disease, on the one hand, and to minimize the adverse economic impact. On the other hand, the lockouts, while they may be effective in reducing new infections, increase the economic gap, damaging the jobs and incomes of tens of thousands of people.

Also in this context, Ashraf (2020) reaffirm that the government's extensive and strict actions, as tight measures of social distancing, aggressive testing and quarantine policy, and generous government income support programs, can reduce the rate of new infections. Abebe (2020) points out that almost all pandemic diseases exhibit their own patterns, which need to be defined by the level of transmission and coverage. Following the outbreak, Coronavirus (COVID-19) has a fast-transmitting nature and is growing exponentially worldwide. Moreover, Livadiotis (2020) considers that normally, the Coronavirus spread evolution curve starts with a pre-exponential phase, which is characterized by a slight logarithmic growth, followed by the outbreak, which is, the exponential growth phase.

Therefore, the search for alternatives on trying to control the pandemic is important to guarantee effectiveness in facing this reality. Chakraborty and Maity (2020) highlight that COVID-19 is a global threat that requires a response involving all countries. Governments should be responsible for providing accurate information to help the public cope with this new infection. To lessen the damage related to COVID-19, public health and infection control actions are needed immediately to limit the global spread of the virus.

2.2 THE PANDEMIC IN BRAZIL

The Brazilian social diversities and their continental amplitude are points to be considered in the analysis for the formulation of strategies to face the pandemic. About that, Coelho *et al*. (2020) point out that Brazil presents strong spatial heterogeneity in terms of demographics, age distribution, access to public health care, and poverty rates. Due to that inequalities, the pandemic of COVID-19 must impact these populations in different ways, if factors such as transmissibility, lethality, and vulnerability are taken into consideration.Therefore, in line with the above, Dantas et al. (2020) evidence that the number of COVID-19 cases continued to grow exponentially due to difficulties in establishing truths and effective social distancing. In the Brazilian real context, a large number of informal workers are still working normally and there is a lack of access to information for a great part of the population about minimal infection prevention and control measures, including hand washing and respiratory etiquette.

The strategies for controlling the spread of COVID-19 were established through the Ministry of Education (MS), having published on June 18, 2020, to the Official Gazette of the Union (DOU), the ordinance establishing general guidelines aimed at the prevention, control, and mitigation of COVID-19 transmission,and the promotion of physical and mental health of the Brazilian population, in order to contribute with actions for the safe resumption of activities and safe social coexistence, (BRASIL, Administrative Rule 1.565, 2020).The orientations are also aimed at promoting the population's physical and mental health. The objective is to support local strategies for a safe resumption of activities and social coexistence, respecting the specificities and characteristics of each sector or branch of activity. It will be up to the local authorities and local health agencies to decide, after evaluating the epidemiological scenario and the response capacity of the health care network, on the resumption of activities.

In addition, the ordinance 1565 adds that among the measures indicated by the MH are the non-pharmacological ones, such as social distancing, respiratory etiquette and hand hygiene, use of masks, cleaning and disinfection of environments, and home isolation of suspected and confirmed cases, which should be used in an integrated way in order to prevent illness and control transmission of COVID-19, it also allows the gradual resumption of activities developed by the many different sectors and a safe return to social life.

The realization of analysis about the design of measures to combat the pandemic exposed at the national level, becomes important according to Garcia and Duarte (2020), the protection of public health must be guiding the decisions to be taken by the managers. It is fundamental that these decisions are based on the best available evidence and communicated in a transparent manner, to promote the confidence of the population. The guidelines of the authorities and the acceptance of the people to NPI (non-pharmacological interventions) will be determinant for the course of the epidemic of COVID-19 in Brazil.

2.3 THE PANDEMIC IN RIO GRANDE DO SUL AND THE STRATEGIES FOR DEALING WITH THE DISEASE

Given the circumstances of the Coronavirus situation, the state government of Rio Grande do Sul has implemented measures to combat the spread of the disease. Therefore, the SES/RS (2020) has defined an action and contingency plan with the task of investigating, managing, and notifying potentially suspected cases of Coronavirus infection. In addition, the state government has defined strategic axes for tackling COVID-19, these are: Health axis, economic axis, social axis, and security axis.

Figure 1 - Strategic axes for facing the pandemic

Source: Government of the State of Rio Grande do Sul (2020). COVID-19 Strategy - Presentation.

Figure 1 presents the strategic axes defined by the Government of Rio Grande do Sul.In the health axis, we highlight the high volume of investments, focused on the structure (expanding the capacity of care, given the demand generated by the Coronavirus), human resources (hiring servers in strategic areas, as well as providing PPE), inputs (sanitation materials and equipment), and monitoring (controlling data, performing tests, and using geolocation).

In the economic axis, as part of the constant monitoring of the cash flow and the market, the government examines the behavior of the gaucho economy since the beginning of the pandemic and uses projections for decision making (RIO GRANDE DO SUL, 2020). Moreover, to mitigate the effects of the measures adopted to fight the pandemic, the government is working on a survey of the impacts by the productive sector to ensure support for the businesses and the maintenance of the services that entrepreneurs need throughout their journey (RIO GRANDE DO SUL, 2020).

In the social scope, a structure was developed to collect and distribute food and hospital supplies, also aiming at prioritizing services via digital services, social assistance to vulnerable groups and temporary suspension of in-person school activities, implementing methodologies according to the different realities in the state. In the security axis, strategic actions of the Secretariat of Public Security (SPS) and the Secretariat of Penitentiary Administration (Sepena) were implemented (RIO GRANDE DO SUL, 2020).

Moreover, State Decree nº 55.128 of March 19, 2020 was instituted, declaring a state of public calamity in the entire territory of the State of Rio Grande do Sul for the prevention and confrontation of the Coronavirus epidemic (RIO GRANDE DO SUL, 2020). In this same perspective, the state government established Decree nº 55.129 of March 19, 2020, establishing the Crisis Cabinet for Coping with the Epidemic COVID-19, Crisis Council for Coping with the Epidemic COVID-19, Inter-institutional Group for Monitoring the Actions for Prevention and Mitigation of the effects of COVID-19 in the Rio Grande do Sul State Prison System and Emergency Operation Center - COVID 19 (EOC COVID-19) of Rio Grande do Sul State.

State Decree nº 55.240 of May 10, 2020 institutes the Controlled Distancing System for the purpose of preventing and confronting the epidemic caused by the new Coronavirus (COVID-19) in the state of Rio Grande do Sul, reiterates the declaration of a state of public calamity throughout the state and makes other provisions. In addition, the decree aimed to implement the flag system to control the risks linked to the state's regional subdivisions and determined the mandatory use of a face shield whenever one is in a collective enclosure, understood as a place intended for permanent simultaneous use by several people, closed or open, private or public, as well as in its circulation areas, on public roads and on means of transport (RIO GRANDE DO SUL, 2020).

It should also be noted that article 7 of State Decree nº 55.240 of May 10, 2020 determines that the disclosure of the results of the measurement of the indicators will occur weekly, always on Saturdays, and the Final Flag in which each Region is classified will be in force from midnight of the Monday immediately after until twenty-four hours of the following Sunday (RIO GRANDE DO SUL, 2020).

2.3.1 Controlled distancing model in the state of Rio Grande do Sul

The controlled distancing model was a strategic measure adopted by the State of Rio Grande do Sul in order to minimize the impacts of the new Coronavirus. To define the flags adopted in the 20 different zones of the state, the methodology is based on the fundamental dilemma present in the controlled opening decision: health versus economic impact. Each sector of economic activity differs in two characteristics: risk of contagion and economic relevance. In a controlled relaxation of the quarantine, the natural candidates to be the first to open are the sectors with low risk of contagion and high economic relevance. By the same logic, the sectors that should remain closed for a longer period of time are those that have high associated risk and low economic impact. What is less clear is what to do with sectors that have low risk but low economic relevance, or high risk but high economic relevance. The decision about which of these activities the restrictions should be relaxed for primarily involves the preferences of *policymaker* about how much more risk one is willing to incur in exchange for more economic activity and vice-versa.

According to the second item of the State Decree nº 55.240, of May 10, 2020, the twenty Regions, corresponding to the grouping of the thirty Health Regions, named after the Municipality with the largest population, are as follows, informed through table 1.

Table 1 - Municipalities with the largest population and corresponding Health Region

Municipality	**Health Region**
Santa Maria	R01 e R02
Uruguaiana	R03
Capão da Canoa	R04 e R05
Taquara	R06
Novo Hamburgo	R07
Canoas	R08
Porto Alegre	R09 e R10
Santo Ângelo	R11
Cruz Alta	R12
Ijuí	R13
Santa Rosa	R14
Palmeira das Missões	R15 e R20
Erechim	R16
Passo Fundo	R17, R18 e R19
Pelotas	R21
Bagé	R22
Caxias do Sul	R23, R 24, R25 e R26
Cachoeirinha do Sul	R27
Santa Cruz do Sul	R28
Lajeado	R29 e R30

Source: Technical Note on the Sectoral Index for Controlled Distancing (2020).

There are seven macro-regions in the state (corresponding to the seven health macro-regions) and twenty regions (corresponding to one or more of the thirty health regions). These are evaluated every two weeks by a committee established by the government. The indicators measure both the spread of the virus and the hospital infrastructure. According to the values of these indicators, one of four "flags" (yellow, orange, red and black) is assigned to each region of the state, which determines the types of suspended activities. Still, for Stein, Sulzbach andLazzari (2020), The model requires that the flags serve as 'alerts', generating lower or higher restrictions depending on the spread of the disease and the health system's capacity to provide care.

2.4 INTERNATIONAL EMPIRICAL LITERATURE

Considering the procedures found in the empirical literature it is possible to divide its scope into three strands, although self-contained, that have very specific discriminated characteristics such as modeling the contagion process and the implications of policies on the topic, research that addresses the effects of measures, and work focused on the use of artificial intelligence and machine learning on the topic.

2.4.1. Modeling the contagion process and policy implications

One of the most relevant works about the counting process is due to Avery *et al*. (2020) which critically reviews the literature from many scientific fields. The same work provides a fundamental basis for economists and other social scientists in conducting original research on the topic.

The authors explain how the Susceptible Infections Recovered-SIR model, just SIR from this point on, works as an epidemiological model. To this end, the authors describe the system of equations and the mechanics of the dynamic process for the solution considering various heterogeneity patterns.

In the same perspective, Atkenson (2020) attempted to adopt an economic perspective to the SIR model applied to the COVID-19 propagation data in the United States for a period between 12 and 18 months. The model simulation pointed to the need for the tightening of social distancing measures for a period between one year and 18 months. Although such results are expressive and similar to the measures adopted in Rio Grande do Sul, the results are based on the counterfactual of social distancing and the use of masks.

2.4.2. Effects of measures against COVID-19

Research based on large databases has been made available. Initially, Stock (2020) schematized the use of the SIR model with goals related to the role of economists in understanding the effects of social distancing and containment policies in relation to pandemic advancement, all of these centered on the interaction with economic variables. As a fundamental result, the author attributed to the parameter of the "asymptomatic rate" the key role in determining the trade-off between the economic cost of the pandemic and the cost of lost lives due to exhaustion of the healthcare system.

Countermanche *et al*. (2020) have analyzed the potential danger of an exponential evolution of the virus contamination caused by the absence of interventions. The method used by the authors was based on the difference-in-differences model for a panel database involving United States counties and time units. The results pointed to the need to provide credible information for the adoption of an adequate strategy regarding the recovery of economic activity.

Hsiang *et al*. (2020) suggested that the benefits of actions taken by authorities in order to reduce the growth of the infection rate could not be directly observed and could be understood only through simulation-based processes. The conclusion was based on the use of reduced-form econometric techniques on data from 1717 sites, regions, and countries related to non-pharmaceutical interventions designated for pandemic containment in locations in China, South Korea, Italy, Iran, France, and the United States.

Chernozhuzov, Kasahara e Schrimpf (2021) presented a comprehensive study involving the combination and ordering of policies to combat the spread of the Coronavirus. The method used was based on machine learning applied to a causal structure that discriminated as: the direct effect of the chosen policy; the effect of the policy adopted through the behavior of agents; and, the effect of direct behavior in response to new information. The specifications were structured in the SIR model for the binary variable cases and deaths with the lags with respect to the three causal effects indicated. From a complete structure, the model allows the conclusion of at least five basic results applied to the United States, namely:

(i) that the new information of virus transmission disseminated by the media, associated with new policies are the main determinants of new cases and deaths;

(ii) changes in government policy explain a large part of the changes observed in the behavior of the population in relation to social withdrawal;

(iii) national decrees regarding the mandatory use of masks could have reduced the growth of the infection rate and weekly deaths by more than 10% by the end of April, and could have reduced the number of deaths by between 19 and 47% by the end of May;

(iv) the absence of decrees regarding mandatory social isolation caused the number of cases to increase between 6 and 63%, and without the closure of commerce the cases could have increased between 17 and 71%; and

(v) uncertainty about the effects of school closures is due to the absence of scanning data across the United States.

Finally, the order of priority would be between wearing masks, closing down commerce and the obligation to stay at home.

Even though such results are expressive, they do not abandon counterfactualism and distance the assumptions from an entirely real world. This point influenced the design of the present work significantly through the motivation to use machine learning applied to a real database.

2.4.3. Artificial intelligence and machine learning in COVID-19 modeling

One of the reasons for using artificial intelligence and machine learning is in Alimadadi *et al*. (2020). The authors considered the use of classifications using artificial neural networks for large database observations for COVID-19 patients related to the breathing patterns of such patients. Such models have become extremely useful for determining common patterns among agents with similar behavior.

Yadav, Perumal and Srinivas (2020) predicted the spread of the virus among the regions of China, the United States, Italy, South Korea, and India. The authors used a comparison between the Support Vector Regression (SVR) model, the Support Vecor Machines (SVM) model, linear regression models, and polynomial regression models. The authors used supervised learning for the period between January 22, 2020 and April 24 of the same year. For a period of 93 days, the authors predicted a catastrophe foretold due to the absence of severe measures to contain the pandemic.

Finally, the model of Ardabili *et al*. (2020) criticizes the lack of accuracy for long-term predictions of more traditional statistical models compared to using the machine learning approach for predicting COVID-19. For alternatives to the SIR and SEIR (Susceptible Exposed Infection Removed) models, the authors considered the incubation period.

The authors have identified many machine learning alternatives such as Random Forest, Neural Networks, Bayesian Networks, Genetic Algorithms, Classification and Regression Trees. In order to highlight the potential use of machine learning, the models were applied to data from Germany, China, Iran, Italy, and the United States for the total number of cases over the 30-day period. For such databases, the authors tested eight alternative specifications, namely logistic, linear, logarithmic, quadratic, cubic, composite, power, and exponential. Unlike traditional statistical models, machine learning models have shown ability in the long-term predictive process.

3 METHODOLOGY: MACHINE LEARNING APPROACH

The methodological approach chosen by the present work is limited to Machine Learning, ML from this point on. Such a choice is due to the broad scope of the literature when data analysis is aimed at the description and prediction of complex structures. In these structures the causal variables present alternatives such that the design has high dimensionality and the out-of-sample prediction process is relevant.

The perspective of using ML in economics has thus provided important tools, and these have had an important impact on economics by being more closely associated with the discipline of Econometrics. To Mullainathan e Spiess (2017), the approach proves adept at designing complex out-of-sample structures. ML, according to Varian (2014), is a subfield of Artificial Intelligence and presents a path for modeling complex relationships with tasks divided into the branches of supervised jobs and unsupervised jobs.

In the present work, the objective is to investigate the factors that explain the status of the evolution of COVID-19 patients in the State of Rio Grande do Sul, between 02/29/2020 and 07/29/2020. This comprises a set of 66473 observations for the outcomes of:

$$\text{evolution} = \begin{cases} 1, & \text{RECOVERED} \\ 2, & \text{DEATH} \\ 3, & \text{IN RECOVERY} \end{cases} \tag{1}$$

For the approach, two *dummies* were specified. The first was initiated on March 19, 2020 from Decree nº 55.129, which established Crisis Cabinet for Coping with the COVID-19 Epidemic, Crisis Council for Coping with the COVID-19 Epidemic, Inter-institutional Monitoring Group for Prevention and Mitigation Actions in the Prison System of the State of Rio Grande do Sul and the Emergency Operation Center - COVID 19 (COE COVID-19) of the State of Rio Grande do Sul (RIO GRANDE DO SUL, 2020). Such subsample (called subsample_1) comprises 66366 observations; the second subsample (called subsample_2) refers to the period beginning May 10, 2020 with the publication of State Decree nº 55.240, establishing the Controlled Distance System for the purpose of preventing and confronting the epidemic caused by COVID-19 at the state level, reiterates the declaration of a state of public calamity in the entire territory, and makes other provisions. It included the establishment of a flag system for the control of risks linked to the regional subdivisions of the state and determined the mandatory use of a face mask whenever in a collective enclosure, understood as a place intended for permanent simultaneous use by several people, closed or open, private or public, as well as its circulation areas, public roads and means of transportation.

With the model for explaining evolution indicated by equation (1), the following analytical frameworks were studied:

- LASSO[1] for Linear Regression
- LASSO for Multinomial Logistic Regression
- Elastic Net Regression
- Regression Tree
- Random Forest

3.1 ANALYTICAL APPROACHES

The use of alternative approaches allows the understanding between them without distancing from the focal point of the work, which is to show the main variables that explain the occurrence of COVID-19 in Rio Grande do Sul.

3.1.1. LASSO of Linear Regression

To use the LASSO or Absolute Minimal Reduction and Selection Operator approach, according to Kleinberg *et al*. (2015) specifies a function that minimizes the mean square error of prediction and penalizes the specification of the function that creates variance. Therefore, algebraically, one can verify that for an Ordinary Least Squares equation:

$$\hat{f}_{OLS} = arg \min_{f_\beta \in F_{linear}} \sum_{=1}^{n} \left(y_i\text{-}f(x_i)\right)^2 \tag{2}$$

[1] LASSO (Least Absolute Shrinkage and Selection Operator) – Absolute Minimum Shrinkage and Selection Operator.

Given equation (3), in a linear specification context, the two key parameters that refer to the mean square error (squared bias) and the variance penalty are denoted by the following two plots:

$$\underbrace{\left(E_D[\hat{y}_0]\text{-}y\right)^2}_{square\ bias} + \underbrace{E_D\left[\hat{f}(x)\text{-}E_D\left[f_{y_0}\right]^2\right]}_{variance} \quad (3)$$

The specification of a framework for ML[2] would be reformulated into a framework where a regularizer that performs a penalty and further parameterized a trade-off between variance and bias. Algebraically, a structure would be specified according to equation (4):

$$\hat{f}_{LM} = arg \min_{f_\beta \in F} \sum_{=1}^{n} \left(y_i\text{-}f(x_i)\right)^2 \quad + \lambda R(f) \quad (4)$$

Wherein:

$R\left(f_\beta\right) = /\!/\beta/\!/^1$ = regularizer that penalizes the specified function that creates variance; and, λ = Lagrange multiplier that balances variance and bias.

In a LASSO specification for a logistic regression, the linear function would be replaced by the multinomial logistic format specification.

Considering the alternatives between Regression Ridge, not contemplated in the present work, Regression LASSO and the Elastic Net Regression. Algebraically, the structure form would be:

$$J(w)_{Ridge} = \sum_{i=1}^{n} \left(y_i\text{-}\hat{y}_i\right)^2 + \lambda \|w\|_2^2,\ \ L2: \lambda \|w\|_2^2 = \lambda \sum_{j=1}^{m} w_j^2 \quad (5)$$

For the generic LASSO specification, analogous to equation (6), one has:

$$J(w)_{LASSO} = \sum_{i=1}^{n} \left(y_i\text{-}\hat{y}_i\right)^2 + \lambda \|w\|_1, \quad L1: \lambda \|w\|_1 = \lambda \sum_{j=1}^{m} \left|w_j\right| \quad (6)$$

What can make the LASSO specification as extremely useful is the fact that some weights "*w*" can become equal to zero.

[2] ML - Machine Learning.

According to Raschka and Mirjalili (2019) the Elastic Net Regression approach would include a linear combination of the Ridge and LASSO Regressions.

3.1.2. Elastic Net Regression

Considering the specification of equation (7) below, the combination between the Ridge Regression and the LASSO Regression, determining a parameter that represents this combination would determine the robustness of the Elastic Net Regression.

$$J(w)_{ElasticNet} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 + \lambda_1 \sum_{j=1}^{m} w_j + \lambda_2 \sum_{j=1}^{m} \left|w_j\right| \tag{7}$$

According to Hasti, Tibshiran and Friedman (2017), the approach selects variables such as LASSO and would do dimensionality reduction together with the predictor coefficients such as the Regression Ridge approach. Also, specifying a penalty would involve:

$$L_q = \lambda \sum_{j=1}^{p} \left(\alpha\beta_j^2 + (1-\alpha)\left|\beta_j\right|\right) \tag{8}$$

Thus, the determination of the parameter α would regulate the combination between the approaches and the respective penalty.

3.1.3. Regression Tree

The use of Regression Trees is recommended in highly complex models where the informational structure combines multinomial variables such as evolution, hospitalization, gender, age group, race or color, and dates, for example. The extent of complexity means that the need for downsizing or regularity makes it possible to determine a causal relationship with high robustness.

The advantages of using a Regression Tree is that induction makes it possible:

(i) better performance on large data sets;

(ii) possibilities for non-linearities and interactions;

(iii) useful in a classification procedure with multiple outcomes or with continuous dependent variables; and,

(iv) non-normal distributions with long tails.

Among the disadvantages are the problems of overfitting the data and poor predictions when new data is used. Nevertheless, the use makes it possible to classify potentially relevant variables. A classic example is the classification for Titanic survivors, found in Varian (2014).

In the case studied here, there are 66.473 outputs, which can present one of three alternatives, consisting of *p* inputs. For the tree, there is the specification of M-partitions or regions $(R_1, \cdots, R_M)$. The answer represents a constant c_m for each region. According to Hasti, Tibshiran and Friedman (2017), the specification would be given by equation (9):

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m) \tag{9}$$

Assuming the smallest sum of the mean square error as the criterion, the tree would be split for the variable *j* in the point *s*. Such divisions into branches would thus obey the optimization criterion given by:

$$\min_{j,\, s}\left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \tag{10}$$

The procedure is subject to the restriction given by relation (11), below:

$$\hat{c}_1 = média\left(y_i | x_i \in R_1(j,s)\right) \text{ and } \hat{c}_2 = média\left(y_i | x_i \in R_2(j,s)\right) \tag{11}$$

At the end of the process of interactions and subdivisions of the tree, several subtrees will be identified. A subtree $T \subset T_0$ can be obtained after pruning the tree on the branch T_0. Therefore, classifying the terminal nodes as *m*, they would represent the region R_m.

A unique way to constrain the tree would be to set a complexity cost as a criterion or penalty. In such a way, the goal would involve the determination of a parameter α, where the subtree $T_\alpha \subseteq T_0$ to minimize $C_\propto(T)$. The analytical structure corresponds to:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha|T| \tag{12}$$

In which:

$N_m = \#\{x_i \in R_m\}$;

$$Q_m(T) = \frac{1}{N_m}\sum_{x_i \in R_m}\left(y_i - \hat{c}_m\right)^2; \text{ and,}$$

$$\hat{c}_m = \frac{1}{N_m}\sum_{x_i \in R_m} y_i$$

In possession of the structure of a tree it is possible to glimpse its main determining elements and the fundamental variables that determined it.

3.1.4. Random Forest

According to Raschka and Mirjalili (2019), In general, a random forest is a set of decision trees, where the aggregation of the trees would represent a forest. To this end, the average of multiple trees, which individually suffer from the problem of high volatility, would make it possible to build the model of a sufficiently robust structure that is less susceptible to the overfitting problem.

In general, the proposal for an algorithm would involve at least four fundamental steps, namely:

(i) in the first step a random sample of size *n* generated via bootstrap without replacement would be drawn;

(ii) a tree would be drawn from the sample. Therefore, at each node you would have

a. would be randomly selected from aspects of the structure without replacement; and,

b. the node would be a divider from the aspects with the goal of maximizing the informational gain of a given objective function.

(iii) steps (i) and (ii) would be repeated k-times; and,

(iv) the prediction from each tree would be aggregated to determine a class called the majority rule.

Although it presents the difficulty of not presenting values for the hyperparameters of the relationship, the approach would present the advantage of allowing the ranking of the determinants of the relationship in order of magnitude.

3.2 DATA AND PROCESSING

To explain the three alternatives of the variable EVOLUTION, the following explanatory variables and covariates that would causally influence it were considered:

- *codibge*[3]: code in integer numbers;

[3] See site: https://ti.saude.rs.gov.br/covid19/

- *dataconfir:* date transformed into value by the Excel spreadsheet;
- *datasintomas:* date transformed into value by the Excel spreadsheet;
- *dataevoluc:* 0 if recovered and date of death transformed into value by Excel spreadsheet;
- *hospital:* binary: 1 = NO; and, 2 = YES;
- *dataobit:* 0 if recovered and date of death transformed into value by Excel spreadsheet;
- *dataevolest:*date transformed into value by the Excel spreadsheet;
- *gender*: assigned the binary values: 0 = male; e 1 = female;
- *age group:* multinomial values: 1 = less than 1 year; 2 = between 1 and 4 years ; 3 = between 5 and 9 years; 4 = between 10 and 14 years; 5 = between 15 and 19 years; 6 = between 20 and 29 years; 7 = between 30 and 39 years; 8 = between 40 and 49 years; 9 = between 50 and 59 years; 10 = between 60 and 69 years; 11 = between 70 and 79 years; and 12 = over 80 years old (inclusive);
- *race/color*: attributes: 1 = WHITE; 2 = BLACK; 3 = DARK; and 4 = NOT INFORMED;
- *region:* see table (1);
- *criterion:* attributes*: 1 = RT-PCR; 2 = FAST TEST; and, 3 = CLINICAL EPIDEMIOLOGICAL.*

The variables chosen as direct explanatory variables were *gender*, *age group*, *race/color*, *region* and *criterion*. The set of covariates is formed by the variables *codibge, dataconfir*, *datasintomas*, *dataevoluc*, *hospital*, *dataobit* and *dataevolest*.

Two alternative procedures were considered in the analysis framework, namely:

(i) the first one divided the sample into ten parts and used the first one for estimation and the others for training until the final results were obtained; and,

(ii) the second, divided the sample into three parts and used the first for estimation and the others for training until the end of the sample.

4 RESULTS AND DISCUSSION

The results were analyzed in blocks, and then they were analyzed together to explain what happened between March 9 and July 29 of the current year.

When including the set of explanatory variables, no specification was obtained for the logistic regression, as per equation (3). So, it was decided to consider the variables alone in the multinomial logistic LASSO regression. Then, the two treatments were idealized for the explanatory variables conforming the following two tables:

Table 2 - LASSO regression variable selection + OLS # treatment 1

Min	-7.73E-01	-7.66E-01	-7.64E-01	-7.66E-01	-7.66E-01
1Q	-2.51E-02	-2.66E-02	–2,630e-05	-2.73E-02	-2.52E-02
Median	3.05E-03	4.47E-03	3.03E-03	4.42E-03	1.64E-03
3Q	2.70E-02	2.62E-02	2.67E-02	2.55E-02	2.55E-02
Max	1.55E+00	1.53E+00	1.54E+00	1.54E+00	1.54E+00
Intercepto	1,24E+00***	1,24E+00***	1,24E+00***	1,24E+00***	1,24E+00***
	(8,83E-06)	(2,79E-05)	(1,18E-05)	(1,43E-05)	(2,01E-05)
datasintomas	1,21E-03***	1,21E-03***	1,21E-03***	1,21E-03***	1,21E-03***
	(9,22E-06)	(9,25E-06)	(9,50E-06)	(9,41E-06)	(9,25E-06)
dataevoluc	-5,50E-01***	-5,50E-01***	-5,50E-01***	-5,50E-01***	-5,50E-01***
	(1,01E-05)	(1,03E-05)	(1,01E-05)	(1,01E-05)	(1,03E-05)
dataobito	1,66E-01	1,66E-01***	1,66E-01***	1,66E-01***	1,66E-01***
	(7,82E-06)	(7,88E-06)	(7,87E-06)	(7,83E-06)	(7,81E-06)
dataevolest	-7,83E-01	-7,83E-01***	-7,83E-01***	-7,83E-01***	-7,83E-01***
	(1,82E-05)	(1,82E-05)	(1,82E-05)	(1,83E-05)	(1,83E-05)
d1dataevoluc	NA	NA	NA	NA	NA
d1dataobito	NA	NA	NA	NA	NA
d1dataevolest	NA	NA	NA	NA	NA
d2dataevolest	3,60E-05	3,78E-05*	3,69E-05*	3,62E-05*	4,11E-05**
	(1,55E-05)	(1,55E-05)	(1,55E-05)	(1,55E-05)	(1,56E-05)
gender	1,63E-05				
	(1,29E-05)				
age group		3,66E-06			
		(3,57E-06)			
race/color			3,14E-06		
			(6,49E-06)		
region				8,58E-07	
				(1,17E-06)	
criterion					-2,53E-05*
					(1,24E-05)
Standard Error Res.:	0,000156	0,0001561	0,0001562	0,0001562	0,0001557
Graus de lib.	587	587	587	587	587
R2 Multip	1	1	1	1	1
R2 Ajusted	1	1	1	1	1
Est-F	1.73E+12	1.73E+12	1.73E+12	1.73E+12	1.74E+12
p-value	< 2,2e-16	< 2,2e-16	< 2,2e-16	< 2,2e-16	< 2,2e-16

Source: Results obtained by the authors. Significance: '***' 0,01%, '**' 0,05 e '*' 0,1.

When considering treatment 1, see table 2, the variables *Intercepto*, *datasintomas*, *dataevoluc*, *dataobito* and *dataevolest* were shown to be relevant in determining the evolution of cases with *dataevoluc* and *dataevolest* showing a negative sign. Such results were analogous in the conventional regression approach, shown in Table 1. Consequently, what can be inferred is that the evolution of treatment, after identification, points to the recovery of the patient, since the value assigned for this result is 1, to the detriment of 2 for death and 3 for the recovering state.

Finally, only the criterion influenced the results, considering the estimation of the equation, noting that criteria such as rapid or clinical epidemiological testing would be related to the case of patient recovery situation. Also, the only variable *dummy* for the second decree, showed a contrary influence to the case of the variable without *dummy* for *dataevolest*.

Now, taking into account the second treatment that infers on a larger data set and only two blocks for training, as shown in table 3, one would have:

Table 3 - LASSO regression variable selection + OLS # treatment 2

Min	-0,04729	-0,04696	-0,04986	-0,04776	-0,04692
1Q	-0,00245	-0,00248	-0,00248	-0,00241	-0,00250
Median	-0,00102	-0,00102	-0,00098	-0,00090	-0,00105
3Q	0,00116	0,00119	0,00131	0,00118	0,00123
Max	194,500	194,540	194,469	194,516	194,536
Intercepto	1,2412658***	1,24E+00***	1,239081***	1,239236***	1,241501**
	(0,000885)	(2,74E-03)	(0,001165)	(0,001404)	(0,002031)
dataconfirm	0,0012436	1,27E-03	0,001488	0,001332	0,001348
	(0,00183)	(1,83E-03)	(0,001831)	(0,001829)	(0,001844)
datasintomas	0,0024385	2,41E-03	0,00241	0,002552	0,002317
	(0,001802)	(1,80E-03)	(0,0018)	(0,001803)	(0,001825)
dataevoluc	-0,5581527***	-5,58E-01***	-0,55821***	-0,55812***	-0,55822***
	(0,000999)	(1,00E-03)	(0,000998)	(0,000999)	(0,001014)
dataobito	0,1736381***	1,74E-01***	0,173671***	0,17366***	0,173633***
	(0,000761)	(7,65E-04)	(0,000761)	(0,000761)	(0,000761)
dataevolest	-0,7794533***	-7,80E-01***	-0,7795***	-0,77912***	-0,77956***
	(0,005818)	(5,82E-03)	(0,005816)	(0,005823)	(0,005824)
d1dataevoluc	NA	NA	NA	NA	NA
d1dataobito	NA	NA	NA	NA	NA
d1dataevolest	-0,001485	-1,44E-03	-0,00143	-0,00174	-0,00139
	(0,005708)	(5,71E-03)	(0,005706)	(0,00571)	(0,005712)
d2dataevolest	-0,0018479	-1,85E-03	-0,00181	-0,00191	-0,00184
	(0,001537)	(1,54E-03)	(0,001537)	(0,001537)	(0,001538)
gender	-0,0006319				
	(0,001269)				
age group		8,76E-05			
		(3,48E-04)			
race/color			0,00135*		
			(0,000704)		
region				0,000161	
				(0,000117)	
criterion					-0,00036
					(0,001266)
Standard Error Res.:	0,04438	0,04439	0,04437	0,04438	0,04439
Graus de lib.	4913	4913	4913	4913	4913
R2 Multip	0,9949	0,9949	0,9949	0,9949	0,9949
R2 Ajusted	0,9949	0,9949	0,9949	0,9949	0,9949
Est-F	1.20E+08	1.20E+08	1,2e+05	1,2e+05	1.20E+08
p-value	< 2,2e-16	< 2,2e-16	< 2,2e-16	< 2,2e-16	< 2,2e-16

Source: Results obtained by the authors. Significance: '***' 0,01%, '**' 0,05 e '*' 0,1.

In detriment to the results in table 2, for the first treatment, as shown in table 3, the variable *datasintomas* proved to be irrelevant for the estimation. The other variables *Intercepto*, *dataevoluc*, *dataobito* and *dataevolest* showed the same results, including the signs. It was noted that the alternation of the explanatory variables had no influence on the results, except for the variable *race/color*. The first block, linked to the *gender* variable, outlined a positive result for the date *evoluc* and *dataevolest*. This points to a trend toward status in treatment, or death throughout the sample studied.

Table 4 - LASSO for Multinomial Logistic Regression - Treatment 1

	Evolution = 1		**Evolution = 2**		**Evolution = 3**	
Variable\lambda	0,01327	0,0000959	0,01327	0,0000959	0,01327	0,0000959
(Intercept)	3,357,942	7,475,175	-2,71E+00	-5,50E+00	-0,64498	-197,912
gender	.	.	.	.	.	-144,827
race/color	.	.	.	2.12E-01	.	.
region	.	-0.02601	.	.	.	0,002527
Codibge	.	.	.	5,60E-01	.	.
datasintomas	.	.	.	-4,03E-01	0,029277	3,644,167
dataevoluc	.	0,034998	.	.	-104,147	-170,526
dataobito	-0,20571	-160,076	8,87E-01	1,10E+00	.	.
dataevolest	1,184,522	1,751,252	.	.	-173,056	-410,763
d1dataevoluc	.	0,016463	.	.	-0,33903	-0,68369
d1dataobito	-0,02175	-0,06631	3,96E-14	1,04E-13	.	.
d1dataevolest	.	0,01213	.	.	.	-0,00299
d2dataevoluc	.	0,044512	.	.	.	-106,382
d2dataevolest	.	0,349648	.	.	.	.
noise1	.	.	.	-6,29E-01	.	.
noise3	.	.	.	.	.	0,029929

Source: Results obtained by the authors.

For the prediction two parameters were considered *lambda* alternative: the *lambda-min*, which results in the minimum mean error of *cross*-validation of the estimation process; *lambda-lse*, which shows the most regularized model such that the error is within the minimum standard error.

Given these alternatives, according to table 4, the following results were obtained for the states of recovery, death, and in recovery:

(i) for recovery, when the *lambda* is given by the minimum mean error of *cross*-validation, the variables *Intercepto*, *dataobito*, *dataevolest* and *d1dataobito* (reflects the weighting of the first decree) determine such events; when the lambda is given by the most regularized model the variable region was included, *dataevoluc*, *d1dataevoluc* (reflects the weighting of the first decree), *d1dataevolest*, *d2dataevoluc* and *d2dataevolest* represented the best model. Such a model emphasizes the role of decrements in the recovery process of infected patients;

(ii) the occurrences of deaths were determined by the intercept, the variable *racacor*, the *codibge*, *dataobito*, *d1dataobito*, and one random variable. In other words, only after the second decree was the situation softened, with elements linked to race and region as predominant factors in a more regularized model; and,

(iii) finally, the occurrences for the recovering state the intercept, the *gender*, *region*, *datasintomas*, *dataevoluc*, *dataevolest*, *d1dataevoluc*, *d1dataevolest*, *d2dataevoluc* (these showing the influence of both decrees) and a certain degree of randomness, measured by the variable *noise3*, enabled the identification of a regularized model.

The patient states presented resulted in differentiated analytical frameworks[4] for the evolution in recovery, death, and in recovery, namely: considering the most regularized models, according to the region of origin, the recovery acted negatively, that is, according to the region, the possibility of not being recovered increases almost in the same proportion as the occurrences; the occurrence of death, on the other hand, is five times lower than the recovery state; when the decrees were considered, it was noted that the former caused the occurrence of death to be proportionally lower.

In the state of death, randomization negatively influenced the occurrence of such a state with symptom identification up to 1.5 times lower in treatment 1.

Considering the state in recovery, the date of symptoms made the state in recovery up to 38 times smaller, which makes recovery itself possible, or even death possible. The estimated evolution date indicated that the state in recovery was up to approximately 61 times lower, making either recovery or death possible. Finally, the effect of the first decree made the chances of such a state up to twice as high, which corroborates the positive effect of such measures.

Table 5 - LASSO for Multinomial Logistic Regression - Treatment 2

	Evolution = 1		**Evolution = 2**		**Evolution = 3**	
Variable\lambda	0,010927	0,000808	0,010927	0,000808	0,010927	0,000808
(Intercept)	3,462,852	5,034,082	-276,706	-392,289	-0,69579	-11,112
Datasintomas	.	.	.	.	0,06746	0,471157
Dataevoluc	.	.	.	.	-0,98095	-222,663
Dataobito	-0,18121	-100,159	1,011,725	0,997672	.	.
Dataevolest	118,928	1,426,449	.	.	-186,741	-336,462
d1dataevoluc	.	.	.	.	-0,52727	-0,60832
d1dataevolest	.	.	.	.	.	-0,16888

Source: Results obtained by the authors.

[4] Consider the values K1 = -2.608, exp (K1) = 0.0736. For this case, the inversion is considered to have a value equal to 13.59 times smaller. For positive values like K2=0.6741, exp (K2) = 1.96; the odds double for the occurrence of the event.

Considering the transformed values of the identified parameters, and because it is a structure divided into three blocks, one for estimation and two for treatment, fewer variables were observed to influence the results.

The analytical values pointed to the dating of the deaths positively influencing the recovery process. This result is due to the fear of the occurrence of death and a greater care by the system for the cases and the opening of relevant structural procedures for the contingency of the pandemic in Rio Grande do Sul.

It can be seen from table 5 that the date of evolution pointed out that it is up to nine times less likely to occur for the recovering state. Analogously, *dataevolest* indicated up to twenty-nine times less chance of such a state occurring.

Table 6 - Elastic Net Regression Parameters

	Treatment 1	**Treatment 2**
(Intercept)	1,245,836	1,243,682
dataconfir	0,005842	0,005105
datasintomas	0,009593	0,007898
dataevoluc	-0,25331	-0,27368
dataobito	0,085547	0,081778
dataevolest	-0,49854	-0,49063
d1dataconfir	0,005558	0,008948
d1datasintomas	0,00262	0,004028
d1dataevoluc	-0,22427	-0,23301
d1dataobito	0,070536	0,082862
d1dataevolest	-0,21372	-0,2404
d2dataobito	.	0,000941
d2dataevolest	-0,01203	-0,01006

Source: Results obtained by the authors.

For the Elastic Net modeling, as shown in table 6, it was observed that the explanatory variables did not show representativeness in any of the treatments. Although the variables and hyperparameters were almost the same, except for the *d2dataobito* present in treatment 2. The justification is that, because it is a larger sample in the second treatment, the date of death weighted by the *dummy*, for the second period, showed an important relative significance in determining the cases of evolution.

Variables such as *dataevoluc*, *dataevolest*, *d1dataevoluc*, *d1dataevolest* and *d2dataevolest* presented negative signs.

Table 7 - Branches of Tree Regression

		Treatment 1				Treatment 2			
	nsplit	**CP**	**error rel**	**xerror**	**xstd**	**CP**	**error rel**	**xerror**	**Xstd**
1	0	0,8313	1,0000	1,0003	0,0303	0,8144	1,0000	1,0001	0,0194
2	1	0,0585	0,1687	0,1688	0,0134	0,0635	0,1856	0,1856	0,0090
3	2	0,0285	0,1102	0,1103	0,0131	0,0333	0,1220	0,1221	0,0088
4	5	0,0096	0,0127	0,0127	0,0039	0,0102	0,0132	0,0133	0,0026
5	6	0,0001	0,0031	0,0036	0,0013	0,0021	0,0030	0,0032	0,0007
6	7					0,0001	0,0009	0,0010	0,0004

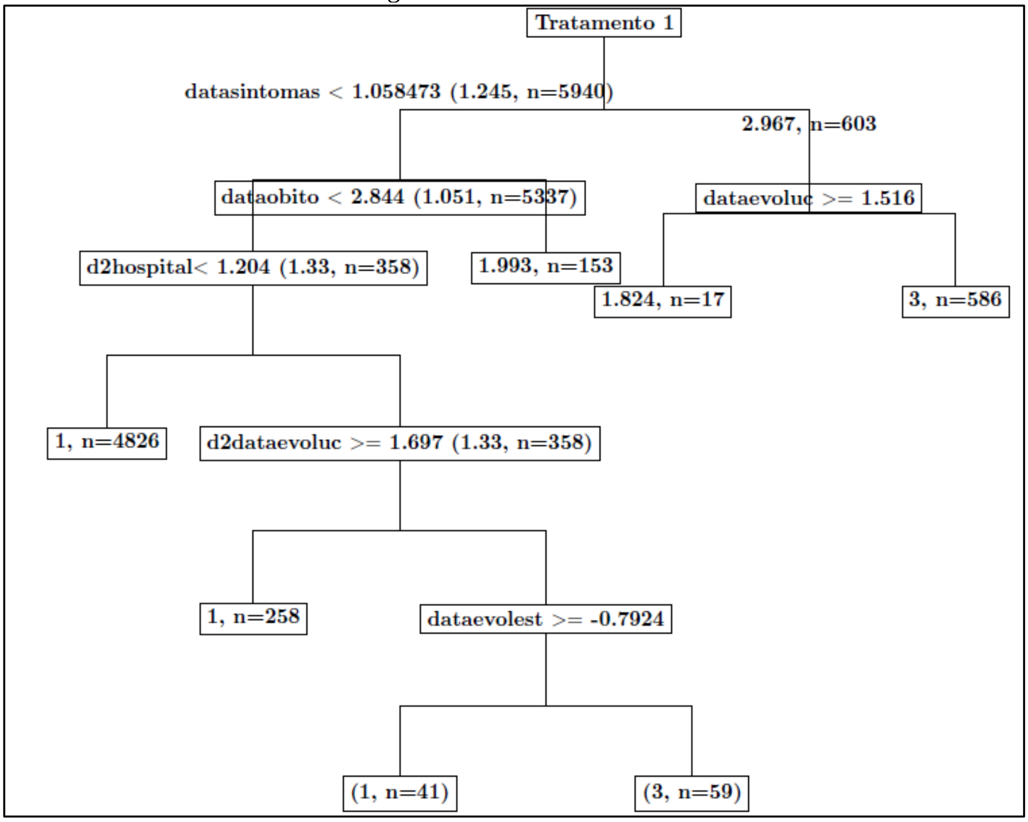
Source: Results obtained by the authors.

The *nsplit* column refers to the number of splits or branches that the regression establishes for the tree throughout the estimation process; the column CP refers to the importance of each partition or each branch; the column *erro rel* refers to the relative error in each partitioning; the column *xerror* dedicates the statistic to the relative value of the estimated error; and the column *xstd* to the standard error of the estimated error.

Then, based on Table 7, the first treatment showed up to six branches outlined in Figure 2. Also, the error becomes minimal when it reaches up to six branches.

The results represent the probabilistic upper bounds of the identified variables in a tree. Graphically, the description of Treatment 1 (Tratamento 1) can be represented by the Figure below:

Figure 2 - Treatment 1 Tree

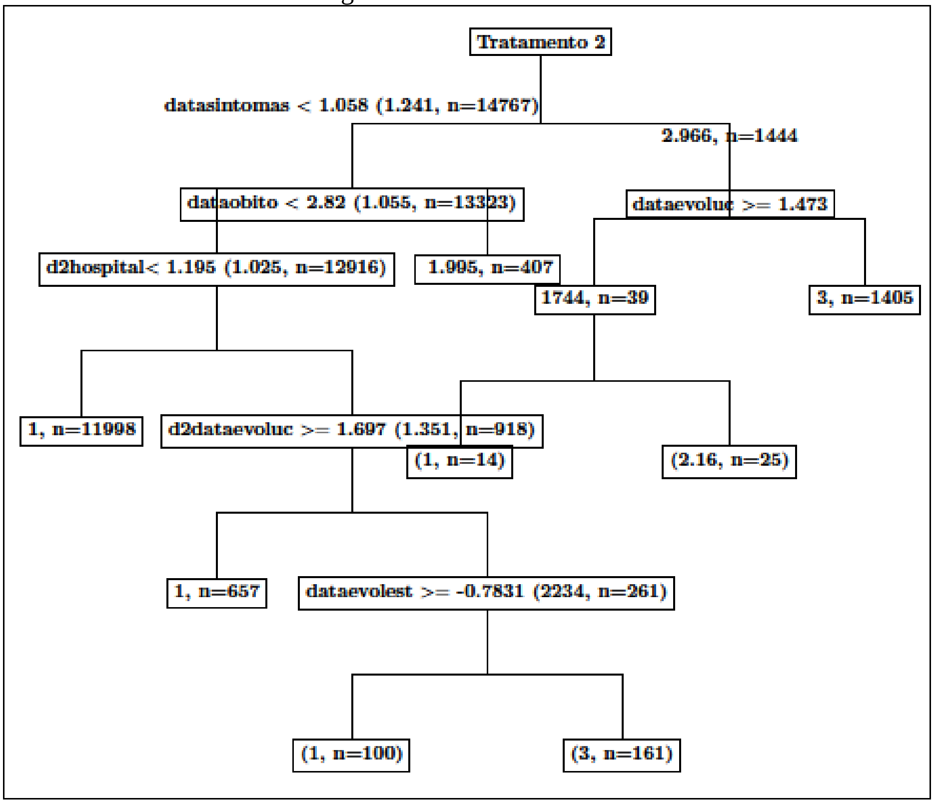


Source: Results obtained by the authors.

The tree for treatment 1 comprised just over six thousand observations for the first partition of the data. Therefore, the tree was divided into six branches with *dataevoluc* equal to or greater than 1.058 pointed to *dataevoluc* as the determinant branch on the right. In this branch approximately 17 deaths and 586 patients in the process of treatment were identified. The branch below 1.058, to the left on the first branch, eventually generated four other branches determined by *dataobito*, *d2hospital*, *d2dataevoluc*, *dataevolest*. With *dataobito* higher or equal to 2.844, 153 deaths were identified, while for a date less than 2.844, 5.337 patients were identified; of the latter, with *d2hospital* with a date less than 1.204 were identified 4.826 recovered and with a date of 1.204 or more, 358 patients were identified. Subsequently, of these patients with *dataevolest* less than -0.7924 evidenced 258 recovered. For *dataevolest* equal to or higher than -0.7924 we had 41 recovered and another 59 in recovery.

It was observed that the structures are very similar, with only the *d1dataevoluc* variable showing, in treatment 1, a probability threshold of up to 1 percentage point. The relative structure is the same for more than ninety-five percent of the trees drawn, even if the final sketch is slightly different.

Figure 3 - Treatment 2 Tree



Source: Results obtained by the authors.

For the second Treatment, as shown in Figure 3, which represents more than sixteen thousand observations, the tree gained an extra branch on the right. Therefore, in the branch on the right, with *datasintomas* higher or equal to 1.058 and then with *dataevoluc* higher or equal to 1.473, 1.045 patients were identified in a state of recovery. With values less than 1.473, 39 patients were identified, with 14 recovered and 25 deaths. Considering, the branching to the left with *datasintomas* less than 1.058, 14.767 patients were identified. Of these, 407 deaths were identified with *dataobito* higher than 2.82. For *dataobito* less than or equal to 2.82, which do not represent deaths because the date is null, 12.916 patients were identified. With *d2hospital* less than 1.195 we observed 11.998 retrieved and 918 of which with *d2evoluc* greater than or equal to 1.697 indicated 918 patients. Such a branch enabled the identification of 657 retrieved. With *dataevolest* equal to or greater than -0.7831, 100 recovered and 161 in recovery were identified.

Table 8 - Random Forest

	Treatment 1				Treatment 2			
ntree	**OOB**	**1**	**2**	**3**	**OOB**	**1**	**2**	**3**
10	0,14%	0,04%	1,88%	0,47%	0,06%	0,03%	0,24%	0,26%
20	0,07%	0,04%	0,00%	0,31%	0,03%	0,02%	0,00%	0,13%
30	0,07%	0,02%	0,62%	0,31%	0,03%	0,02%	0,00%	0,13%
40	0,05%	0,02%	0,00%	0,31%	0,02%	0,02%	0,00%	0,06%
50	0,05%	0,02%	0,00%	0,31%	0,03%	0,02%	0,23%	0,06%

Source: Results obtained by the authors.

Considering table 8, it was noted that Treatment 1 showed a low and stable error rate of around 0.02%; for the death cases, above 1 stabilized only after the 40th tree. Finally, the processes of evolution to the recovering state showed little sensitivity to the increase in the number of trees in the forest, inferring that only after the second decree was it possible to identify such results with greater accuracy.

For Treatment 2, again, recovered cases were stabilized after the inclusion of the block and twenty trees. The deaths were stabilized, but after the fortieth tree the results became more inaccurate, which required verification of the confusion matrix.

Table 9 - Random Forest confusion matrices

	Treatment 1					Treatment 2			
	1	**2**	**3**	**Classification errors**		**1**	**2**	**3**	**Classification errors**
1	5130	1	0	0,000194894	**1**	12769	2	0	0,000156605
2	0	162	0	0,000000000	**2**	1	425	0	0,002347418
3	0	2	645	0,003091190	**3**	0	1	1569	0,000636943

Source: Results obtained by the authors.

Analyzing the random forest confusion matrix, as shown in Table 9, it was noted that Treatment 1 classified up to 50 trees with up to 5 variables in the structure divisions and with an estimated error rate of 0.02%. Also, it showed only three false positives (negatives) with one recovered falsely classified as death and two in recovery also classified as death.

From an analogous perspective, Treatment 2 classified up to 50 trees with up to five variables in the structure divisions and with an estimated error rate of 0.03%. For this treatment there were four false positives (negatives) with two recovered individuals falsely classified as deaths, one death misclassified as recovered, and one recovering falsely classified as death.

It is notable that the model enables accurate classification to the detriment of other alternatives in the econometric literature, which shows the importance of such classification processes.

Thus, instead of using counterfactualism, it was observed that the use of machine learning, even though it shows high long-term predictive power, as proposed by Aradabili *et al*. (2020), shows good accuracy in the causal explanation of the determinants of the states of the patients with COVID-19, for the analyzed period in Rio Grande do Sul.

5 CONCLUSION

The current pandemic scenario made it possible to study the impacts of the coping measures adopted by the state of Rio Grande do Sul in the development of the number of confirmed cases, the number of deaths, and also the number of patients under treatment. Due to the lack of specific work on the policies adopted in Rio Grande do Sul, the present work used the machine learning approach for the study.

It was decided to construct two blocks, each of which represents the influence of the decrees. For this, the effects on two variables were also considered *dummies*. Hence, for both treatments the inclusion of *dummies* pointed to significant results in the approach explaining the systematic learning process in relation to the event. In addition, the identification of a variable *noise* of the events made it possible to infer some degree of randomness.

Contrary to the use of counterfactual approach, the present work used different structures for each of the states (recovered, death, in recovery) were identified in the use of LASSO Regression variable selection. Accordingly, when the decrees were considered, it was noted that the former caused the occurrence of death to be proportionally lower. Furthermore, for the state of death, randomness negatively influenced the occurrence of such a state.

For treatment 1, a large effect of immediacy was observed on the results with the effect of the first decree on the chances of such a state up to two times greater, corroborating the positive effect of such measures. From a political point of view, the fear of the occurrence of death and a greater care of the system for the cases and the opening of procedures structure relevant to the contingency of the pandemic in Rio Grande do Sul.

For the Elastic Net modeling, it was observed that the explanatory variables (Gender, Age Group, Race, Region, Criterion) showed no representativeness in any of the treatments. However, the LASSO Regression showed more significant results as the value is closer to zero from the left. From a more general point of view, the second treatment showed more significant results for the LASSO Regression.

For the Tree and Random Forest Regression procedures, the variables showed the following relevance: *d1datasintomas*, *d2datasintomas* and *datasintomas* presenting up to 23% relevance each in the study; *d1dataconfir*, *d2dataconfir* e *dataconfir* presenting up to 5% each; *dataevolest*, *d2dataevoluc*, *d1dataobito*, *dataobito*, *dataevoluc*, *d2dataobito*, *d1dataevolest* and *d2dataevolest* presenting up to 2% each and *d1dataevoluc* presenting up to 1% relevance.

The number of trees in the forest in the random forest procedure made possible structures of up to fifty trees with few results from the twentieth tree and excellent accuracy.

To corroborate the adequacy and providence of the measures adopted by the government of Rio Grande do Sul, Figueiredo, Polli and Andrade (2020) studied problems related to the number of cases based on data from the PNAD-COVID-19 postulating that the actual number of infected is much higher than reported by testing due to two possible biases related to low testing numbers and false-negative results from rapid tests. Therefore, the authors aimed to use the method *semi-bayesiano* proposed by Wu *et al*. (2020) for the correction of the indicated biases. They concluded that the existence of pre-existing immunity and the edge of collective immunity made it possible to reduce contamination and control it. For Rio Grande do Sul they reported that the contamination rate was less than 1% and with the correction by the method *semi-bayesiano*, This would be the result of the adequacy of the policies adopted by the government. This would be the result of the adequacy of the policies adopted by the government.

Although none of the works in the international literature cited highlight the characteristics of the agents in the evolutionary process of the patients, they are limited to counterfactualism and not to the factual evidence of the approach presented here.Furthermore, as a research agenda, it is suggested that the same approach adopted for the second wave data be used, which would allow inferences to be made about the behavior of agents in relation to the first wave measures and the effects of explanatory variables and their covariates.

It was concluded that the decrees had positive and long-lasting effects on the effects of the pandemic in Rio Grande do Sul, and the use of a new database would corroborate the estimates made up to July 29 of this year.

REFERENCES

ABEBE, T. H. (2020), Forecasting the number of Coronavirus (COVID-19) cases in Ethiopia using exponential smoothing times series model. **MedRxiv**, p. 1-12.

AGGARWAL, C. C. (2018), Neural Networks and Deep Learning: A Textbook, Cham (Switzerland): **Springer International Publishing**.

AL-ROHAIMI, A. H.; AL OTAIBI, F. (2020), Novel SARS-CoV-2 outbreak and COVID-19 disease; a systemic review on the global pandemic. **Genes & Diseases**, p. 1-22.

ALIMADADI, A.; ARYAL, S.; MANANDHAR, I.; MUNROE, P.; JOE, B.; CHENG, X. (2020), Artificial intelligence and machine learning to fight COVID-19, **Physiol Genomics**, v.52, p. 200-202.

ARDABILI, S.; MOSAVI, A.; GHAMISI, P.; FERDINAND, F.; VARKONYI-KOCZY, A.; REUTER, U.; RABCZUK, T.; ATKINSON, P. (2020), COVID-19 outbreak prediction with machine learning, **Algorithms**, v.13, n. 249, p. 1-36.

ASHRAF, B. N. (2020), Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets. **Journal of Behavioral and Experimental Finance**, v. 27, p. 1-9.

ATHEY, S. (2018), The impact of machine learning on economics. In: **The economics of artificial intelligence: An agenda**, University of Chicago Press, p. 507-547.

ATKENSON, A. (2020), What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios. **NBER-National Bureau of Economic Research**, Working Paper nº 26867.

AVERY, C.; BOSSERT, W; CLARK, A.; ELLISON, G.; ALISSON, S. (2020), Policy implications of models of the spread of Coronavirus: perspectives an opportunities for economists. **NBER-National Bureau of Economic Research**, Working Paper nº 27007.

BRASIL, Portaria nº 1.565, de 18 de junho de 2020. Estabelece orientações gerais visando à prevenção, ao controle e à mitigação da transmissão da COVID-19, e à promoção da saúde física e mental da população brasileira, de forma a contribuir com as ações para a retomada segura das atividades e o convívio social seguro. **Diário Oficial da União**. Brasília: DF, 2020. Disponível em: <http://www.in.gov.br/en/web/dou/-/portaria-n-1.565-de-18-de-junho-de-2020-262408151>. Acesso em: 28 jul. 2020.

BRASIL. Decreto 30.691, de 29 de março de 1952. Disponível em: Acesso em: 14 jan. 2013.

CHAKRABORTY, I.; MAITY, P. (2020), COVID-19 outbreak: migration, effects on society, global environment and prevention. **Science of the total environment**, p. 1-7.

CHERNOZHUKOV, V.; KASAHARA, H.; SCHRIMPF, P.; (2021), Causal impact of masks, policies, behavior on early COVID-19 pandemic in the US, **Journal of Econometrics**, v. 220, p.23-62.

COURTEMANCHE, C.; GARUCCIO, J.; LE, A.; PINKSTON, J.; YELOWITZ, A. (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate. **Health Affairs**, v. 29, n.7, p.1237-1246.

COELHO, F. C.; LANA, R. M.; CRUZ, O. G.; CODECO, C. T.; VILLELA, D.; BASTOS, L.; GOMES, M. F. (2020), Assessing the potential impact of COVID-19 in Brazil: mobility, morbidity and the burden on the health care system. **MedRxiv**, p. 1-17.

CVETKOVIC, V. M.; NIKOLIĆ, N.; NIKOLIĆ, U. R.; ÖCAL, A.; NOJI, E. K.; ZEČEVIĆ, M. (2020), Preparedness and preventive behaviors for a pandemic disaster caused by COVID-19 in Serbia. **International Journal of Environmental Research and Public Health**, v. 17, n. 11, p. 1-23.

DANTAS, R. C. C.; CAMPOS, P. A.; ROSSI, I.; RIBAS R. M. (2020), Implications of social distancing in Brazil in the COVID-19 pandemic. **Infection Control & Hospital Epidemiology**, p. 1-2.

DUARTE, M. D. Q.; SANTO, M. A. D. S.; LIMA, C. P.; GIORDANI, J. P.; TRENTINI, C. M. (2020), COVID-19 e os impactos na saúde mental: uma amostra do Rio Grande do Sul, Brasil. **Ciência & Saúde Coletiva**, v. 25, p. 3401-3411.

FIGUEIREDO, E.; POLLI, D. A.; ANDRADE, B. B. de. (2020), Relatório de Pesquisa: Estimação dos Níveis de Infecção por COVID-19 no Brasil, **PPOGM-UFPEL** – Working Paper.

GARCIA, L. P.; DUARTE, E. (2020), Intervenções não farmacológicas para o enfrentamento à epidemia da COVID-19 no Brasil. 2020. **Epidemiol Serv. Saúde**, v. 29, n. 2, p. 1-4.

GEMICIOGLU, B.; BÖREKÇI, Ş.; DILEKTAŞLI, A. G.; ULUBAY, G.; AZAP, Ö.;

HALE, T.; PETHERICK, A.; PHILLIPS, T.; WEBSTER, S. (2020), Variation in government responses to COVID-19. **Blavatnik school of government working paper**, v. 31, p. 1-23.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. (2017), The Elements of Statistical Learning: Data Mining, Inference, and Prediction. **Nature Switzerland: Springer**.

HSIANG, S.; ALLEN, D.; ANNAM-PHAN, S.; BELL, K.; BOLLIGER, I.; CHONG, T.; DRUCKENMILLER, H.; HUANG, L.; HULTGREEN, A.; KRASOVICH, E.; LAY, P.; LEE, J.; ROLF, E.; TSENG, J.; WU, T. (2020), The effect of large-scale anti-contagion policies on the COVID-19 pandemic. **Nature**, v. 584, p. 262-267.

KLEINBERG, J.; LUDWIG, J.; MULLAINATHAN, S.; OBERMEYER, Z. (2015), Prediction Policy Problem. **American Economic Review**, v. 105(5), p. 491-495.

LIVADIOTIS, G. (2020), Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19. **Plos one**, v. 15, n. 5, p. 1-21.

MULLAINATHAN, S.; SPIESS, J. (2017), Machine learning: an applied econometric approach. **Journal of Economic Perspectives**, 31(2): 87-106.

PINTO, N. G. M.; CORONEL, D. A.; MÜLLER, A. P. (2020), Uma discussão sobre o Observatório Socioeconômico da COVID-19 por meio de uma perspectiva estadual, regional e nacional. **Research, Society and Development**, v. 9, n. 7, p. 1-15.

RASCHKA, S.; MIRJALILI, V. (2019), Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. **Packt Publishing Ltd**.

RIO GRANDE DO SUL. Decreto Estadual n° 55.128, de 19 de março de 2020. Declara estado de calamidade pública em todo o território do Estado do Rio Grande do Sul para fins de prevenção e de enfrentamento à epidemia causada pelo COVID-19 (novo Coronavírus), e dá outras providências. **Diário Oficial**, Porto Alegre, 2020. Disponível em: < https://saude-admin.rs.gov.br/upload/arquivos/202003/19125910-decreto-55-128-20.pdf>. Acesso em: 28 jul. 2020.

RIO GRANDE DO SUL. Decreto Estadual n° 55.129, de 19 de março de 2020. Institui Gabinete de Crise para o Enfrentamento da Epidemia COVID-19, Conselho de Crise para o Enfrentamento da Epidemia COVID-19, Grupo Interinstitucional de Monitoramento das Ações de Prevenção e Mitigação dos efeitos do COVID-19 no

Sistema Prisional do Estado do Rio Grande do Sul e Centro de Operação de Emergência - COVID 19 (COE COVID-19) do Estado do Rio Grande do Sul. **Diário Oficial**, Porto Alegre, 2020. Disponível em: < https://saude-admin.rs.gov.br/upload/arquivos/202003/20112207-decreto-55-129-20.pdf>. Acesso em: 28 jul. 2020.

RIO GRANDE DO SUL. Decreto nº 55.240, de 10 de maio de 2020. Institui o Sistema de Distanciamento Controlado para fins de prevenção e de enfrentamento à epidemia causada pelo novo Coronavírus (COVID-19) no âmbito do Estado do Rio Grande do Sul, reitera a declaração de estado de calamidade pública em todo o território estadual e dá outras providências. **Diário Oficial**, Porto Alegre, 2020. Disponível em: < https://saude-admin.rs.gov.br/upload/arquivos/202005/12091118-55-240.pdf >. Acesso em: 28 jul. 2020.

RIO GRANDE DO SUL. Estratégia COVID-19 – Apresentação, 2020. Disponível em: <https://estado.rs.gov.br/apresentacao-estrategia-covid-19>. Acesso em: 28 jul. 2020.

SARYAL, S. (2020), Turkish thoracic society expert's consensus report: recommendations for pulmonary function tests during and after COVID 19 pandemic. **Turkish Thoracic Journal**, v. 21, n. 3, p. 193-200.

SECRETARIA ESTADUAL DA SAÚDE DO RIO GRANDE DO SUL (SES/RS). Plano de Contingência e Ação Estadual do Rio Grande do Sul para Infecção Humana pelo novo Coronavírus (2019-nCoV). **Secretaria de Saúde (SES)**, 2020. Disponível em: < https://saude.rs.gov.br/upload/arquivos/202004/11151537-plano-de-acao-corona-2020-rs-versao-10.pdf>. Acesso em: 28 jul. 2020.

SOUSA, G. O.; SALES, B. N.; RODRIGUES, A. M. X.; ROCHA, G. M. M. (2020), Evolução epidemiológica da COVID-19 no Brasil e no mundo. **Research, Society and Development**, v. 9, n. 7, p. 1-13.

STEIN, G.; SULZBACH, V. N.; LAZZARI, M. (2020), Nota Técnica sobre o Índice Setorial para Distanciamento Controlado. **Comitê de Dados do Rio Grande do Sul**, v.1, n. 1, Porto Alegre.

STOCK, J.; (2020), Data gaps, and the policy response to the novel Coronavirus. **NBER-National Bureau of Economic Research**, Working Paper nº 26902.

VARIAN, H. R. (2014), Big data: New tricks for econometrics. **Journal of Economic Perspectives**, 28(2), p.3-28.

WORLD HEALTH ORGANIZATION. 2019 Novel Coronavirus (2019-nCoV): Strategic Preparedness and Response Plan. **WHO**, 2020. Disponível em: <https://www.who.int/publications/i/item/strategic-preparedness-and-response-plan-for-the-new-coronavirus>. Acesso em: 28 jul. 2020.

WU, S. L.; MERTENS, A. N.; CRIDER, Y. S.; NGUYEN, A.; POKPONGKIAT, N. N.; DJAJADI, S.; SETH, A.; HSIANG, M. S.; COLFORD, Jr. J. M.; REINGOLD, A.; ARNOLD, B. F.; HUBBARD, A.; BENJAMIN-CHUNG, J. (2020), Substantial underestimation of SARS-CoV-2 infection in the United States. **Nature Communications**, https://doi.org/10.1038/s41467-020-18272-4.

YADAV, M.; PERUMAL, M.; SRINIVAS, M. (2020), Analysis on novel Coronavirus (COVID-19) using machine learning method. **Chaos, Solitons and Fractals**, v. 139, p.1-12.