A comparative analysis of COVID-19 trends in Colombia for policy makers

Anonymous Authors[1]

1. Introduction

As many countries in the developing world, the Colombian health system is particularly unprepared for a rapid increase in the number of patients with COVID-19. Even though the extraordinary increase in coverage since the nineties when only 24 percent of the Colombian population had health coverage, there were already many barriers to access health care before the pandemic landed in the country.

Several studies have been done to apply machine learning methods to analyze the trends of China, Europe, and the USA (Fong et al., 2020; Ramírez-Hassan & Montoya-Blandón, 2020; Santos et al., 2020); however, there are no models focused on the analysis of Latinoamerican countries that need to take measures to stop the spread of the virus urgently. Current COVID-19 forecasting methods have classified in statistical approaches, stochastic theory approaches, analysis of social media feeds, and data science approaches (Mahalle et al., 2020).

Using mathematical modeling, we can observe that the transmission rate must be lowered to 0.25 to slow the spreading rate(Wu et al., 2020), and it is well known that early intervention policies in public health to mitigate the outbreak of COVID-19 are crucial, yet many countries have failed to implement such measures successfully. Many complex factors interplay in this scenario, such as the economic impact of social distancing, miss-information, the lack of understanding of the functional biology of the virus, and ineffective treatments for critical patients in the Intensive Care Unit.

Here we present an analysis to support policy maker in Colombia. The first step of our pipeline is to identify the behavior of the trend of the virus spread in Colombia using traditional statistical methods, then we compare Colombian trend with the 94 countries that had their first COVID-19 before Colombia using unsupervised machine learning to identify similarities between countries. Once we identify the Countries with the most similar trends, we analyze the policies put in place on those countries and their outcome, to conclude with the best policies to be taken in Colombia at this time.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

2. Methods

Here we present our pipeline to compare virus spread among countries and select best policies for Colombia. Our policy making model, takes as input historic data from all countries, time of policy making, and textual articles describing policy and policy outcome. The output is a set of selected policies for slowing the number of new cases in a country.

First we analyzed historical data per country we performed goodness-of-fit test for exponential distribution for the variables transmission, death, and recovery. The data per country was obtained from the COVID-19 Data Repository by Johns Hopkins CSSE (Dong et al., 2020). Then, we partition the countries using the temporal data, the parameters of the model, Bayesian information criteria (BIC), and their R^2 using traditional hierarchical clustering (Rokach & Maimon, 2005).

Additionally, we extract the dates where five different policies were put in place for 94 different countries using web scraping and topic modeling. We manually curated the results and selected the following five policies and their starting date, closing of fluvial, land and air transport, and stay-at-home order (Ministerio de Relaciones Exteriores). Based on such results we estimate the effectively of the measures as a change of the spread rate.

3. Results

Figure 1 shows the spread curves of Mexico and Peru, the two countries that were found to be the most similar to Colombia's pattern at a cut off of 30 days since the fist positive case. Colombia's curve was fitted to the curve $y = 6.212e^{0.2438x}$ with independent variable x as the day since the first case and y the number of new cases ($R^2 = 0.9802$). Mexico and Peru were modeled to curves $3.5654e^{0.2589x}$ and $5.1382e^{0.241x}$, respectively ($R^2 = 0.9595$, and $R^2 = 0.9143$).

To analyze the different policies ordered, we did not include Mexico and Peru because these countries do not have enough predictive data to include in the forecasting model.

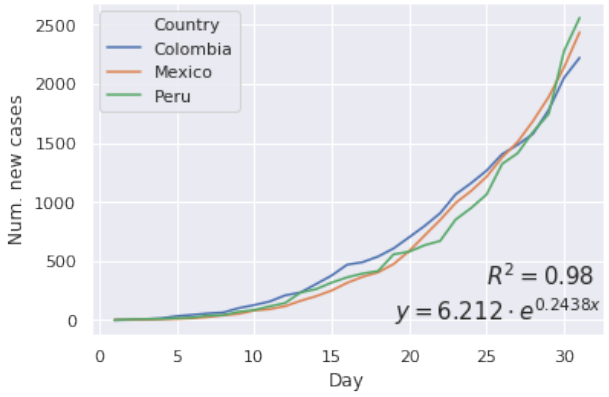

Figure 1. Comparison of the trend of Peru, Mexico and Colombia. Peru and Mexico are two countries that are the most similar to Colombia.

In Figure 2, we compare the extracted dates of the day of starting stay-at-home orders for China, USA (shown only California which is the state with fewer number on cases to date), Spain, Italy, Germany, and France. Our analysis did not evidence a direct correlation between the starting time of stay-at-home orders and the number of new cases for these six countries. We believe that future studies are needed to extract and compare combination of policies and the measures taken to enforce them.

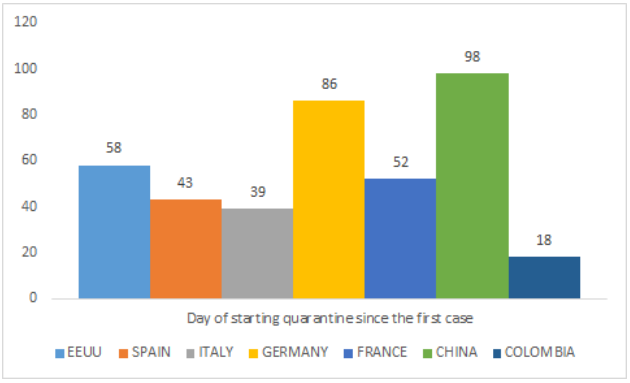

Figure 2. Comparison of the starting date of stay-at-home orders among countries.

4. Conclusions

In the midst of the current COVID-19 pandemic, Colombian officials and general populations are particularly worried about controlling the speed of the virus spreed on the possibility of a collapse of an already weak health system. It is urgent to find the best policies to avoid regrettable outcomes like those seen in Italy, Spain, and Ecuador. To guide our analysis we first use traditional statistical methods to identify the behavior of the spread trend in Colombia and compare it with 97 countries using ML methods to identify similarities. Once we identify the most similar trends, we analyze the policies put in place on those countries and their outcome, to conclude with the best policies to be taken in Colombia at this time. We are incorporating demographics, weather, and other variables to strengthen our model.

References

Dong, E., Du, H., and Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 2020.

Fong, S. J., Li, G., Dey, N., Crespo, R. G., and Herrera-Viedma, E. Composite Monte Carlo Decision Making under High Uncertainty of Novel Coronavirus Epidemic Using Hybridized Deep Learning and Fuzzy Rule Induction. *arXiv preprint arXiv:2003.09868*, 2020.

Mahalle, P., Kalamkar, A. B., Dey, N., Chaki, J., Shinde, G. R., et al. Forecasting models for coronavirus (covid-19): A survey of the state-of-the-art. 2020.

Ministerio de Relaciones Exteriores, C. Medidas adoptadas a nivel mundial ante la pandemia del covid-19. URL minrel.gob.cl.

Ramírez-Hassan, A. and Montoya-Blandón, S. Forecasting from others' experience: Bayesian estimation of the generalized Bass model. *International Journal of Forecasting*, 36(2):442–465, 2020.

Rokach, L. and Maimon, O. Clustering methods. In *Data mining and knowledge discovery handbook*, pp. 321–352. Springer, 2005.

Santos, M., Mariscal, L. C., Henríquez, B., Garzón, J., González, P., Carrera, J. P., Tello, J., Koo, S., Pascale, J. M., and Burkett-Cadena, N. Implementation of bamboo and monkey-pot traps for the sampling cavity-breeding mosquitoes in Darién, Panama. *Acta Tropica*, 205:105352, 2020.

Wu, J. T., Leung, K., and Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697, 2020.