COVID-19 in Mexico: Identifying health and socioeconomic variables for death and hospitalization prediction.

Roberto Barroso Luque* Oscar E. Noriega Villarreal†
Jesica María Ramírez Toscano‡ Luz Stephanie Ramos Gómez§

DRAFT: June 8, 2020

Abstract

The current pandemic caused by the SARS-CoV-2 (COVID-19) virus represents an unprecedented challenge to the global economy. While herculean efforts have been made to analyze, predict, and model the socioeconomic impacts of the pandemic to guide relevant policy, the majority of this analysis has revolved around just a handful of countries. [RESULTS OVERVIEW].With this project, we hope to shed some light on the public health crises currently unfolding in Mexico and, more importantly, provide data-driven policy recommendations to combat the pandemic.

*barrosoluquer@uchicago.edu
†onoriega@uchicago.edu
‡jramireztoscano@uchicago.edu
§stephanieramos@uchicago.edu

1 Introduction

On January 30th, the World Health Organization (WHO) declared the SARS-CoV-2 outbreak a Public Health Emergency of International Concern, yet despite the institution's warnings, political leaders have failed to act in a timely fashion.

As of June 2020, about 7 million people have been infected, and more than 400,000 have died from the COVID-19 pandemic. Until a vaccine is developed, which experts predict will take more than a year, the most efficient policy to hinder viral transmission is mass social-isolation. Unfortunately, while social isolation measures stem the spread of the virus they also have a devastating economic effect.

Millions of jobs have been lost, children have stopped going to school, entire economies are in standby. Countries, such as Canada and the US have provided emergency funds to help middle- and low-income households through social protection mechanisms. In contrast, developing countries' fiscal response has been timid at best and non-existent at worse. Economies such as Mexico and India, where at least half of the labor population works in the informal sector (meaning no access to social protection and/or financial institutions), face additional obstacles.

In Mexico, the Central Bank anticipates a GDP decline of up to 8.8%[1] in a worst-case-scenario, while private institutions such as Credit Suisse expect a drop of 9.6%[2] in GDP. Given its limited resources, it is of imperative importance that the Mexican government prioritize the most vulnerable areas and allocates resources accordingly. However, due to a lack of research and analysis, it is far from clear which states and municipalities are most at risk.

The purpose of this project is to fill this analysis gap to identify vulnerable communities and speed up the governmental response. To do so, we will rely on a Machine Learning approach for:

- Identifying the best socioeconomic and health predictors for COVID-19 deaths and hospitalizations among positive cases.

- Gaining insights into the most at risk populations and locations based on key pre-

[1] *Quarterly Report, January-March 2020*- Banco de Mexico

[2] *Mexico's GDP will have its worst decline since 1932: Credit Suisse* - Forbes Mexico, 04/30/2020

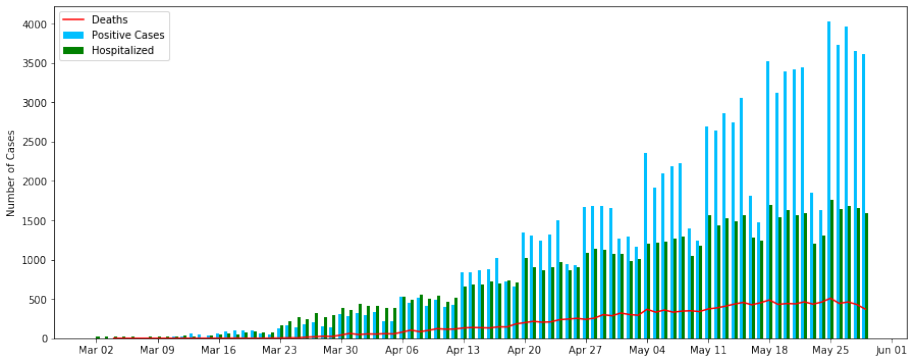
dictors to inform resource allocation.

***

2 COVID-19 in Mexico

Ever since the first COVID-19 officially confirmed case appeared in Mexico City at the end of February[3], as in most countries, the number of both positive cases and deaths has been on the rise with, still, no clear sign of an outbreak curve flattening.

Figure 1: COVID-19 Daily New Cases, Hospitalizations and Deaths in Mexico



Source

3 Data

4 Machine Learning Problem

SUB1

*define classification problem and briefly define models

SUB2

+SMOTE

SUB3

*index building

[3] *Mexico confirms first coronavirus case in the country* - El País, 02/28/2020

5 Main Results

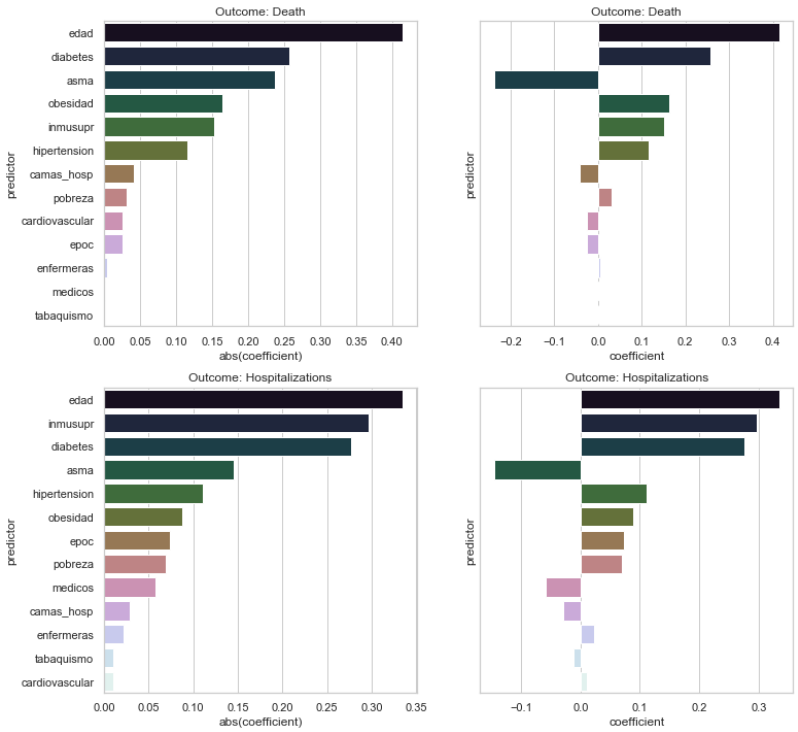
Table 1: Evaluation metrics for COVID-19 related deaths predictions

Prediction outcome: *Death*	Metrics:		
	Recall	*Precision*	*Accuracy*
Linear Support Vector	0.73	0.23	0.71
Logistic Regression	0.71	0.24	0.72
Decision Tree	0.87	0.19	0.58
Balanced Random Forest	0.82	0.21	0.65
Weighted Random Forest	0.77	0.23	0.69
Complement Naive Bayes	0.74	0.15	0.54

Table 2: Evaluation metrics for COVID-19 related hospitalization predictions

Prediction outcome: *Hospitalization*	Metrics:		
	Recall	*Precision*	*Accuracy*
Linear Support Vector	0.82	0.21	0.65
Logistic Regression	0.67	0.56	0.70
Decision Tree	0.88	0.19	0.57
Balanced Random Forest	0.73	0.57	0.71
Weighted Random Forest	0.73	0.57	0.70
Complement Naive Bayes	0.71	0.44	0.59

Figure 2: Feature importance for COVID-19 deaths and hospitalizations prediction using SVM classifier generating synthetic data (SMOTE) per each fold in CV



6 Conclusions

*policy recommendations

SUB1

*Considerations & caveats

7 Appendix

*additional graphs