# Novel Application in Machine Learning: Predicting the Issuance of COVID-19 Stay-at-Home Orders in Africa

*Carter Rhea, Jordan Mansell, and Gregg Murray*

*L'Université de Montréal, University of Western Ontario, Augusta University*

## Covid-19 & SAHOs

Since the Covid-19 Pandemic began over 15 months ago, the lives of citizens around the globe have been disrupted in order to reduce the transmission of the virus. Many governments have opted to put in place **Stay-at-Home-Orders (SAHOs)** which restrict the movement of their citizens by imposing strict guidelines for international, state or provincial, and even city-wide travel. These Stay-at-Home-Orders have a **huge impact on the local population and the country's economy** -- therefore, the implementation and rigidity of these orders have varied strongly from country to country. Moreover, reasons for the issuance of these Stay-at-Home-Orders have been difficult to pinpoint and several sources quote drastically different driving factors in these decisions.

## Problem Statement

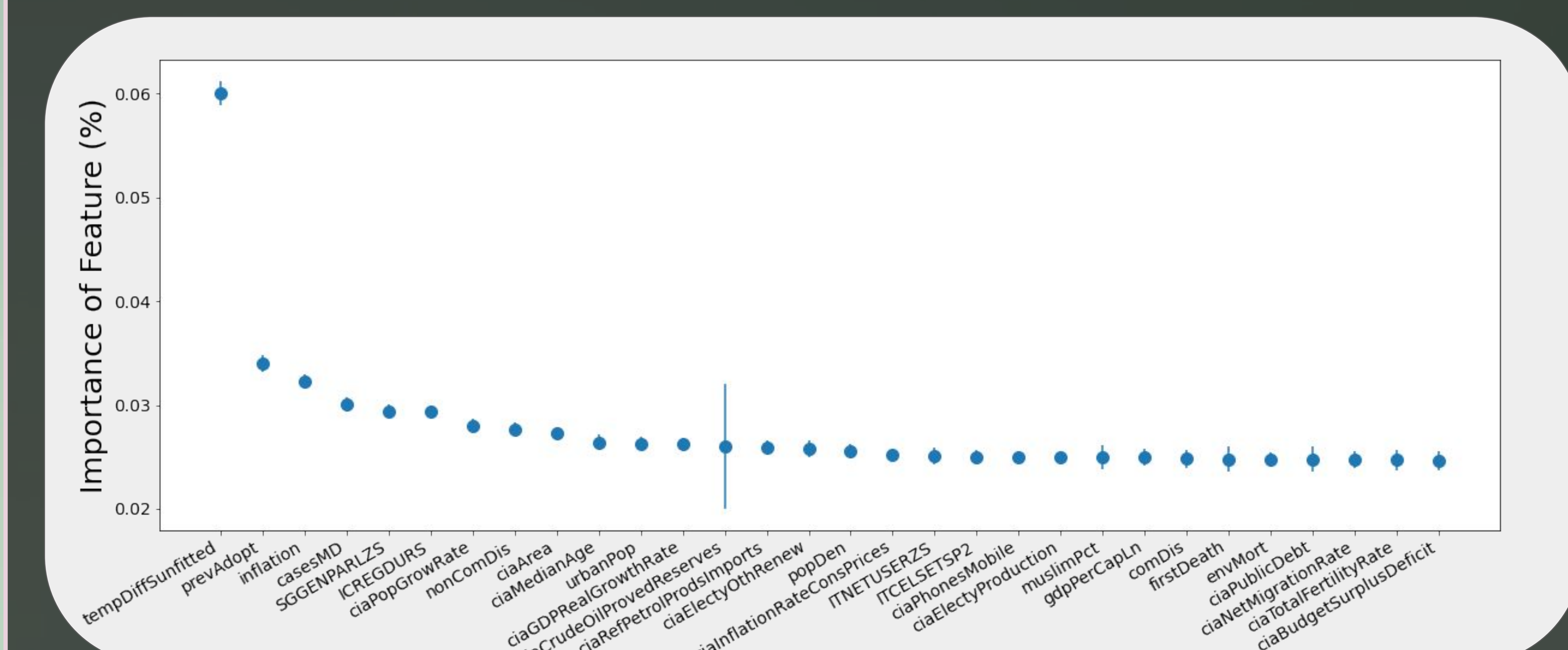
In the work, we investigate which factors are most important for the issuance of SAHO among n=54 African countries between January 31st and June 15th 2020. **We employ a novel dataset of 260 different variables capturing country-level information on economic, political, social, external, and health related factors** for each country in our dataset. To identify the most significant factors, we treat the question of which countries issue SAHO as a classification problem (issued orders vs. did not issue orders) and utilize a random forest classifier, a method of machine learning, to identify the variables that best explain whether a country issued a SAHO order or not.

## Random Forest Classifier Implementation

Our implementation of the random forest uses 50 decision trees. We apply a maximum recursion level of 30 and allow for standard automatic pruning as implemented in sklearn in order to reduce overfitting. We set aside 40 randomly selected countries (~75% of the total sample) to serve as the training set, while the remaining 14 countries (~25%) are used to test the algorithm. Variables described in the problem statement serve as inputs (independent variables) to train the tree while the sahoStatusEHA serves as the target (dependent) variable. To further reduce bias and overfitting, we implement a Monte-Carlo implementation of the random forest algorithm with 1,000 iterations. We calculate the feature importance for each variable from the 1,000 iterations.

| Number of Variables in Final Training | Test Percentage |
|---|---|
| 5 | 68% |
| 10 | 75% |
| 15 | 77% |
| 20 | 70% |
| 25 | 75% |

## Preliminary Results




## Top Variables

1. TempDiffSunfitted - A nonlinear trend variable that is constructed by taking the square root of the number of days between a given day and the day with the highest hazard rate
2. prevAdopt - Percent of African countries previously issuing a SAHO
3. Inflation - Annual inflation rate, 2018
4. CasesMD - Cumulative cases per MD.
5. SGGENPARLZS - Proportion of seats held by women in national parliaments (%)
6. ICREGDURS - Time required to start a business.
7. ciaPopGrowRate - The average annual percent change in the population, resulting from a surplus (or deficit) of births over deaths and the balance of migrants entering and leaving a country.
8. nonComDis - Probability of dying from non-communicable disease, ages 30-70.
9. ciaArea - Sum of all land and water areas
10. ciaMedianAge - Country median population age.

## Conclusions:

We report that our methodology predicts the issuance of SAHO with 75% accuracy. We find 10 variables, defined above, to be crucial to proper categorization. Additionally, we demonstrate that a variety of variables play a critical role in the issuance of Stay-at-Home-Orders rather than a single categorical variable.

## Contact Information:

carter.rhea@umontreal.ca
https://crhea93.github.io