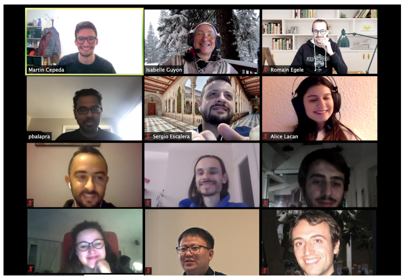
 

TRANSATLANTIC TEAM QUALITATIVE SUBMISSION

Github repo: https://github.com/transatlantic-team/Pandemic-Prize

I. Introduction of the research team and research directions

We are a team of 12 people from 3 continents: Europe, America, and Asia, with our leader (Martin Cepeda) currently based in Paris, France. Our team members include 5 undergraduate students, 1 post-doc, 3 senior researchers, and 3 university professors. The quantitative submission was mostly the work of the undergraduate students. Other team members contributed: infrastructure, data preparation and visualisation, and insight. In terms of infrastructure, we set up a private competition among ourselves on the Codalab platform on which team members made submissions, to monitor progress. We also set up a Github repo to share code.

While, due to time constraints, our team didn't have time to explore many avenues, even if our team members are experienced in machine learning (see e.g. the Google scholar pages of Sergio Escalera, Xavier Baro, Prasanna Balaprakash and Isabelle Guyon) and have studied the prediction and control of the Covid-19 epidemic (Yu et al, 2020; Cepeda, 2020). Hence we hope to make a more significant contribution in the second phase, if we are selected.

We are particularly interested in applying during the Prescriptor phase of the challenge Causal Modeling and Reinforcement Learning (RL) methods and in developing policy optimization, taking into account multiple factors. Since data availability plays a central role in obtaining good models, we started collecting additional data. Our axes for developing our models in Phase II include:

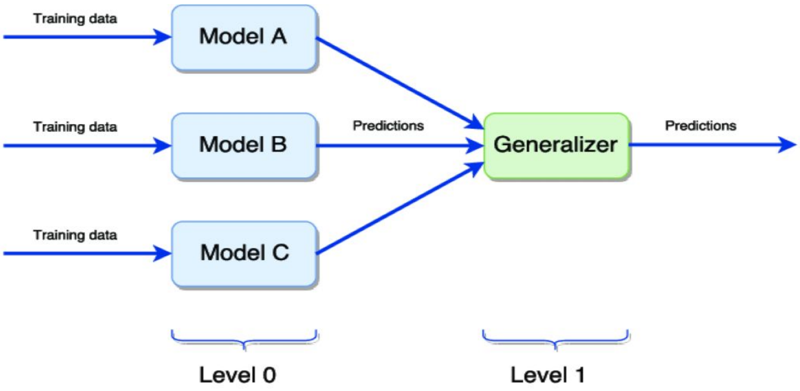
- **Causal analysis:** Policy evaluation using on-line data from previous pandemics remains a more or less open subject due to the novelty of this problem (e.g., Petersen, 2020). A notable example of this subject is a data-driven policy effect forecast (Vanderschaar, 2020) offering a counterfactual analysis framework for different countries. We are aware of the limitations of black-box predictive models based only on observable evolution of a pandemic, and the need to relate predictions to causes (e.g. lockdowns) (Goodman-Bacon, 2020). Structural equation modeling is a promising avenue, complementary to compartimental models (Pearl, 2009).
- **Economic factors:** While most authors according to our research focus on epidemiological models and neglect immediate economic impact, our interest is in blending epidemiology and econometric models, mostly short-term effects such as unemployment, business and school closures, transport reduction, which affect policy-maker decisions. Other indicators such as stock market indices, do not necessarily correlate with the economic impact of COVID-19. However, reduced morbidity and access to good healthcare facilities correlates to GDP growth (Alkire, 2018). More generally an increase in adult morbidity leads to a fall of economic growth (Javaid, 2015). Economic factors can be taken into account by appropriately defining the RL rewards, as done in (Cepeda, 2020).
- **Collateral health impact:** Another important aspect of policy optimization concerns collateral death and collateral adverse effects on public health, including untreated non-Covid related acute or chronic conditions, such as cancer, renal insufficiencies, or depression (Brodeur, 2020). For this reason, estimating secondary effects of untreated conditions due to health facilities being prioritized for COVID-19 in conjunction to pandemic response is important (Aron, 2020; Woolf, 2020). Future work could include creating a Causal/RL framework generalizing classical epidemiology models, which generally don't take into account such effects.

II. Innovation

For our pre-selection submission, our principal innovative point has been to create a meta-model based on several predictors, switching according to countries. We collected additional country-specific time independent data (see section IV). Although the individual models (Lasso, Ridge, SVR) were each trained on all the countries to gain robustness in predictive power (without country-specific data), our meta-model (figure 1) uses country-specific specialisation, making use of country-specific time independent data.

Final submission is about: Predicting 7 days smoothed daily new number of cases with robustness:

Figure 1 (source) Schematic of the submitted meta-model with stacking



- StackingRegressor with RidgeCV as main estimator to regularize
- Sub regressors: RidgeCV, LassoCV, BayesianRidge, LinearSVR

III. Generality

Our model performs rather well across all regions (see detailed results in appendix C.III). The following table summarizes the statistics for MAE per 100K in October test run:

Mean	Std	Median	Min	Max
1.80	4.37	0.23	0.00	32.98

However, we distinguish **3 types of regions** mutually exclusive between them according to the model's performance:

- Type 1: (MAE per 100K < 4) The model performs well across 1 month of predictions (206 out of 236 regions).
- Type 2: (MAE per 100K in [4, 10]) Bosnia and Herzegovina, Nepal, Romania, Portugal, Georgia, Italy, Costa Rica, Botswana, Moldova, Croatia, Peru, Brazil, Ireland, Iceland, Estonia, Poland, Cape Verde, Luxembourg, Oman, Slovenia, Belize, United Kingdom (country).
- Type 3: (MAE per 100K > 10, up to 33) Outliers where the model performs particularly bad: Switzerland, Netherlands, France, Spain Bahamas, Belgium, Czech Republic, Israel and Andorra.

As the regions with the worst performance (Type 3) have in general a rather small population (Andorra, Israel, Bahamas, Czech Republic) the normalized performance per 100K inhabitants is severely penalized. For instance, even when the predictions in Belgium have roughly the same "raw" MAE as in Croatia, the MAE per 100K is 5.7 times bigger in Belgium.

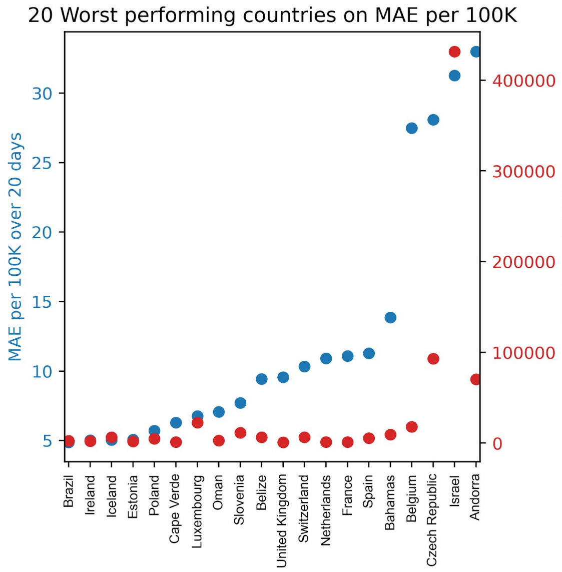


Figure 2 Countries sorted by increasing MAE per 100K. Small countries are among worst performance countries due to normalization by 100K inhabitants

We can see the previous phenomena in figure 2: the gap between raw MAE and MAE per 100K is bigger when the country is smaller. For the Prescriptor phase, we'll pay special attention to the MAE in small countries.

We explain also the good performances in Type 1 and 2 countries mainly because of the stage of the pandemic: those countries have had a monotonous increase in cases and/or haven't reached a wake peak, whereas Type 3 regions are on a second, more powerful pandemic wave or never ended a first one (as in the US), which is a more complex evolution that our model fails to capture.

We show now a sample prediction for countries in each type (20 days prior to prediction period also shown):

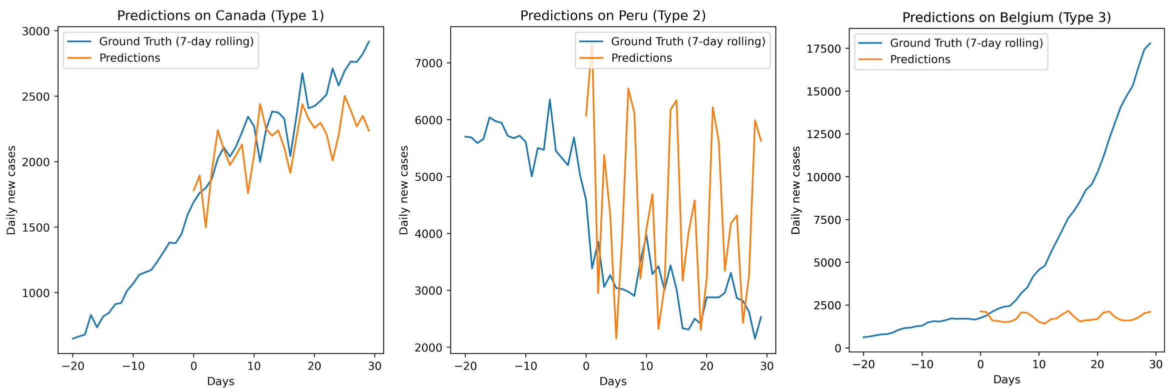


Figure 3 Sample predictions in countries from Types 1, 2 and 3. Negative values on X axis are last days of the training period

Additional figures comparing different models can be found in Appendix C.IV.

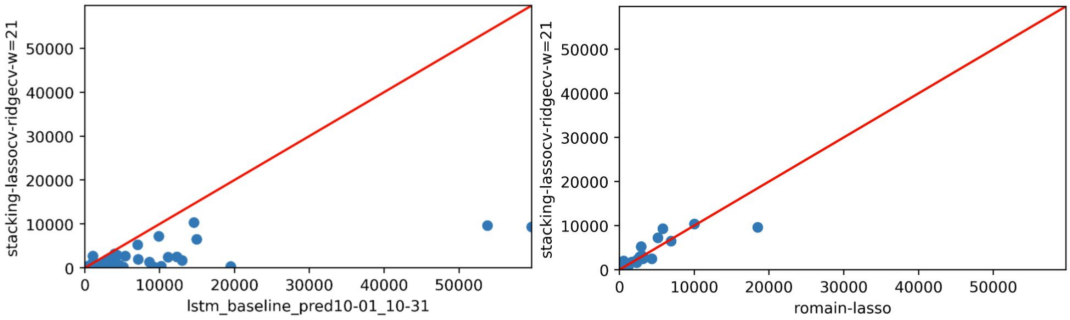
IV. Collaborative contributions

Our code and country-specific time independent data are open-sourced.

V. Consistency

We have homed in on linear predictive models because they seem most consistent on the short and long range. We are aware that this may not seem very refined, however, we compared with a variety of more complex models and found that the simplest models were the most robusts. A detailed analysis between different models can be found in annex C.II. Here we present the performance comparison between our submitter model, the LSTM baseline and a Lasso model:

Figure 4 Performance comparison between submitted model and baselines: Lasso Linear Regression and LSTM



In the plots, each point is a country and the axis represents MAE on 2 different models. The further the points are from the diagonal line (same error in both models), the greater the difference in performance. For instance, our model performs 5 times better than the given baselines over all countries in the period of 1 month of predictions (see annex C.II. for more information on model selection).

VI. Speed and resource use

Our model being based on linear predictors, it largely respects the time constraint imposed by the challenge (predictions in all countries on 180 under 1 hour) both to train and at prediction time: for prediction, running our model over all countries/regions for 180 days takes **less than 4 minutes** in the provided Sandbox environment.

For the submitted model on CPU Intel i7 6 cores: **Training up to 21st of December, 5 mins 30 seconds**. Predicting 30 days for all countries takes 19 seconds.

VII. Addressing the challenge

Prior to start modelling, we explored the available data (see Appendix C.I. and our repo). We discovered that a) new cases time series is very noisy, due to a certain periodicity (more cases are reported on Mondays) and changes in counting methodology per country since the beginning of the pandemic (which results in negative daily new cases), b) NPI data is not consistent (all countries have NaNs) and overall c) data from the first months must be taken cautiously, as COVID-19 was in an early stage and testing, data gathering and individual countries' response was not fully developed.

From this, we chose to rely on a) cumulated number of cases and b) 7-day smoothed new cases, as training input. Also we considered predicting (apart from daily new cases) the daily rate of change in new cases (used as a building block in autoregressive models). See Appendix C.V. for the exploratory modelling. We decided to take into account for the final modelling only the provided data and no external series such as deaths or bed occupancy, because these data would not have been updated during the testing phase. We did not attempt to exploit any loophole in the provided sandbox/predictor API or whatsoever.

VIII. Explanation

The stacking model presented in section II.is trained on the whole OxCGRT dataset (01 of January to 21 of December). As it is an ensemble method (see section II), it first trains the sub-regressors (RidgeCV, LassoCV, BayesianRidge, LinearSVR) and then the meta RidgeCV regressor. Note that we also learn from NPIs as they are part of the dataset.

During the training and prediction stage, we consider a lookback window of 28 days to predict the next day. Predictions over an arbitrary time window (see section III) are computed via a rollout algorithm (Sutton 2018).