Machine Learning (ML) algorithms: Prediction of SARS-CoV-2 cases in Bolivia

Juan Ruiz Otondo[1,2]
[1] Universidad San Francisco Xavier (USFX)
[2] Universidad de Buenos Aires (UBA)

Abbreviated abstract: *The objective of this article is to examine the implications of Machine Learning (ML) algorithms in predictive accuracy of SARS-CoV-2 cases. The methodology is based on the statistics and predictive analytics of ML algorithms. The result determines a global test of 92% in prediction accuracy, the Arima, Randon Forest, Prophet algorithms have a weak test in predicting. For the above reason it could be concluded that the predictive models Glmnet and Prophet W/XGboot errors are better algorithms to predict SARS-CoV-2 case scenarios in the short term.*

Related publications:
– Ruiz, O. Juan. Economic Issues Journal. Central Bank of Ecuador (57), (2021)

Datastudy Consulting

otondojr86@gmail.com

4th Conference on **Statistics and Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Problem, Literary

- **Problem**
  - What are the algorithms that best predict the behaviors of SARS-CoV-2 cases in Bolivia?
  - The objective of this article is to examine the implications of Machine Learning (ML) algorithms in predictive accuracy of SARS-CoV-2 cases.
- **Literary revision**
- Mechanistic models based on susceptible, exposed, infected and recovered states (SEIR) (Bjornstad, Shea, Krzywinski, & Altman,, 2020) which is an extension of the model of susceptible states S (t), infected I (t) and recovered R (t) (SIR);
- Time series prediction models such as ARIMA (Sahai, Rath, Sood, & Singh, 2020), Gray's model (Zhao, Shou, & Wang, 2020) and Markov chain models (Marfak, Achak, Azizi, Nejjari, & Aboudi, 2020);
- The need to use machine learning methods to predict the behavior of cases and deaths due to Covid-19 was concluded (Alballa & Al-Turaiki, 2021)
- Using ML, PROPHET, POLY-MLP algorithms, accurately predicted the number of infected cases per day by training the 25-day sample data (Shahriare, Koushik, Mufti, & al., 2021)
- With ARIMA, RANDON FOREST, BOOSTING (Iwendi, Peshkar, Sujatha, & al., 2020), also with LOGISTIC REGRESSION, ANN, W / XGBOOT ERRORS (Alhayani & Kwekha, 2021)

Datastudy
Consulting

otondojr86@gmail.com

Methods

- It is methodologically based on the mathematical-statistical literature of predictive analytics; ARIMA, RANDON FOREST, PROPHET, GLMNET, PROPHET W / XGBOOS ERROS.
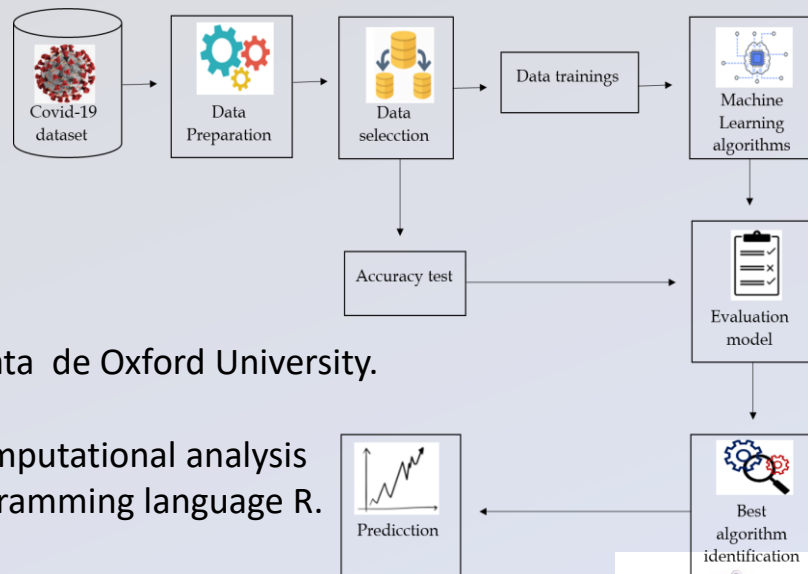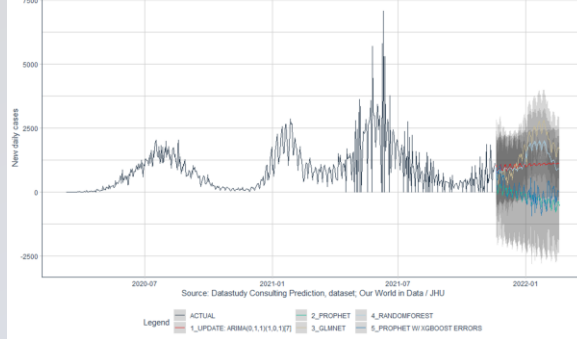
- Data: Johns Hopkins University y Our World in Data de Oxford University.

- To train the Machine Learning algorithms, the computational analysis technique was used with the multiparadigm programming language R.

- The prediction has been running since 01/01/2021, in 30 and 90 days of each prediction.

otondojr86@gmail.com
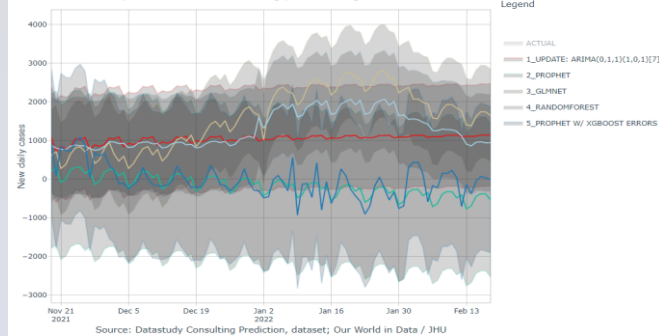
Results and Conclusions
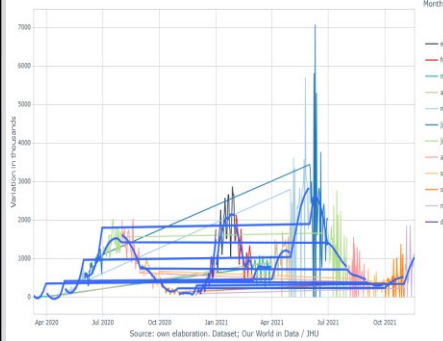
- The Machine Learning algorithms that best predict future contagion cases are PROPHET W / XGBOOS ERROS, GLMNET, followed by PROPHET, to this character is added the weakness of precision of the ARIMA and RANDOM FOREST algorithms, however the global test of predictive precision reaches 91%.
- Given the strong "negative" association between the rigor of the restrictions and the spread of disease, there is no other option to complement the mild restriction policies with awareness, financial assistance and health schemes, before implementing quarantines or restrictions.

Datastudy Consulting

otondojr86@gmail.com